SEQUENCE LISTING

<110> MULLINS, James I.
RODRIGO, Allen G.
LEARN, Gerald H.
LI, Fusheng
NICKLE, David C.
JENSEN, Mark A.

<120> ANCESTRAL AND COT VIRAL SEQUENCES, PROTEINS AND IMMUNOGENIC COMPOSITIONS

<130> 16336-001320US

<140> 10/780,507
<141> 2004-02-17

<150> US 10/204,204
<151> 2001-02-16

<150> PCT/US01/05288
<151> 2001-02-16

<150> US 60/183,659
<151> 2000-02-18

<150> US 60/447,586
<151> 2003-02-14

<160> 125

<170> PatentIn version 3.1

<210> 1
<211> 2652

<212> DNA

<213> Artificial Sequence

<220>

<223> Ancestral HIV-1 group M, subtype B, env sequence

<400> 1

```
atgcgcgtga agggcatccg caagaactac cagcacctgt ggcgctgggg caccatgctg      60
ctggggatgc tgatgatctg ctccgcggcc gagaagctgt gggtgaccgt gtactacggc     120
gtgcccgtgt ggaaggaggc caccaccacc ctgttctgcg ccagcgacgc caaggcttac     180
gacaccgagg tccacaacgt gtgggccacc cacgcctgcg tgcccaccga ccccaacccc     240
caggaggtgg tgctggagaa cgtgaccgag aacttcaaca tgtggaagaa caacatggtg     300
gagcagatgc acgaggacat catcagcctg tgggaccaga gcctgaagcc ctgcgtgaag     360
ttaacccccc tgtgcgtgac cctgaactgc accgacgacc tgcgcaccaa cgccaccaac     420
accaccaaca gcagcgccac caccaacacc accagcagcg gcggcggcac gatggagggc     480
gagaagggcg agatcaagaa ctgcagcttc aacgtgacca ccagcatccg cgacaagatg     540
cagaaggagt acgccctgtt ctacaagctg gacgtggtgc ccatcgacaa cgacaacaac     600
aacaccaaca acaacaccag ctaccgcctc atcaactgca acaccagcgt gatcacccag     660
gcctgcccca aggtgagctt cgagcccatc cccatccact actgcacccc cgccggcttc     720
gccatcctga agtgcaacga caagaagttc aacggcaccg gcccctgcac caacgtgagc     780
accgtgcagt gcacccacgg catccgcccc gtggtgagca cccagctgct gctgaacggc     840
agcctggccg aggaggaggt ggtgatccgc agcgagaact tcaccgacaa cgccaagacc     900
atcatcgtgc agctgaacga gagcgtggag atcaactgca cgcgtcccaa caacaacacc     960
cgcaagagca tccccatcgg ccctggccgc gccctgtacg ccaccggcaa gatcatcggc    1020
gacatccgcc aggcccactg caacctgtcg cgagccaagt ggaacaacac cctgaagcag    1080
atcgtgacca agctgcgcga gcagttcggc aacaacaaga ccaccatcgt gttcaaccag    1140
agcagcggcg gcgaccccga gatcgtgatg cacagcttca actgcggcgg cgaattcttc    1200
tactgcaaca gcacccagct gttcaacagc acctggcact tcaacggcac ctggggcaac    1260
aacaacaccg agcgcagcaa caacgccgcc gacgacaacg acaccatcac cctgccctgc    1320
cgcatcaagc agatcatcaa catgtggcag gaggtgggca aggccatgta cgcccccccc    1380
atcagcggcc agatccgctg cagcagcaac atcaccggcc tgctgctgac tcgagacggc    1440
ggcaacaacg agaacaccaa caacaccgac accgagatct tccgccccgg gggcggcgac    1500
atgcgcgaca actggcgcag cgagctgtac aagtacaagg tggtgaagat cgagcccctg    1560
ggcgtggccc ccaccaaggc caagcgccgc gtggtgcagc gcgagaagcg cgccgtgggc    1620
atgctgggcg ccatgttcct gggcttcctg ggcgccgccg gcagcaccat gggcgccgcc    1680
agcatgaccc tgaccgtgca ggcccgccag ctgctgagcg gcatcgtgca gcagcagaac    1740
```

aacctgctgc gcgccatcga ggcccagcag cacctgctgc agctgaccgt gtggggcatc 1800
aagcagctgc aggcccgcgt gctggccgtg gagcggtacc tgaaggacca gcagctgctg 1860
ggcatctggg gctgcagcgg caagctgatc tgcaccaccg cggtgccctg gaacgccagc 1920
tggagcaaca agagcctgga caagatctgg aacaacatga cctggatgga gtgggagcgc 1980
gagatcgaca actacaccgg cctgatctac accctgatcg aggagagcca gaaccagcag 2040
gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctgtg gaactggttc 2100
gatatcacca actggctgtg gtacatcaag atcttcatca tgatcgtggg cggcctggtg 2160
ggcctgcgca tcgtgttcgc cgtgctgagc atcgtgaacc gcgtgcgcca gggctacagc 2220
cccctgagct tccagacccg cctgcccgcc ccccgcggcc ccgaccgccc cgagggcatc 2280
gaggaggagg gcggcgagcg cgaccgcgac cgcagcgggc gcctggtgaa cggcttcctg 2340
gccctgatct gggacgacct gcgcagcctg tgcctgttca gctaccaccg cctgcgcgac 2400
ctgctgctga tcgtggcccg catcgtggag ctgctgggcc ggcgcggctg ggaggccctg 2460
aagtattggt ggaacctgct gcagtactgg agccaggagc tgaagaacag cgccgtgagc 2520
ctgctgaacg ccaccgccat cgccgtggcc gagggcaccg accgcgtgat cgaggtggtg 2580
cagcgcgcct gccgcgccat cctgcacatc ccccgccgca tccgccaggg cctggagcgc 2640
gccctgctgt ga 2652

<210> 2

<211> 883

<212> PRT

<213> Artificial Sequence

<220>

<223> Ancestral HIV-1 group M, subtype B, env sequence

<400> 2

Met Arg Val Lys Gly Ile Arg Lys Asn Tyr Gln His Leu Trp Arg Trp
1   5   10   15

Gly Thr Met Leu Leu Gly Met Leu Met Ile Cys Ser Ala Ala Glu Lys
            20   25   30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr
        35   40   45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val
    50   55   60

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65   70   75   80

```
Gln Glu Val Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95

Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
            100                 105                 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125

Asn Cys Thr Asp Asp Leu Arg Thr Asn Ala Thr Asn Thr Thr Asn Ser
    130                 135                 140

Ser Ala Thr Thr Asn Thr Thr Ser Ser Gly Gly Gly Thr Met Glu Gly
145                 150                 155                 160

Glu Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn Val Thr Thr Ser Ile
                165                 170                 175

Arg Asp Lys Met Gln Lys Glu Tyr Ala Leu Phe Tyr Lys Leu Asp Val
            180                 185                 190

Val Pro Ile Asp Asn Asp Asn Asn Asn Thr Asn Asn Asn Thr Ser Tyr
        195                 200                 205

Arg Leu Ile Asn Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro Lys
    210                 215                 220

Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Thr Pro Ala Gly Phe
225                 230                 235                 240

Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn Gly Thr Gly Pro Cys
                245                 250                 255

Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val Val
            260                 265                 270

Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Val Val
        275                 280                 285

Ile Arg Ser Glu Asn Phe Thr Asp Asn Ala Lys Thr Ile Ile Val Gln
    290                 295                 300

Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr
305                 310                 315                 320

Arg Lys Ser Ile Pro Ile Gly Pro Gly Arg Ala Leu Tyr Ala Thr Gly
                325                 330                 335

Lys Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Leu Ser Arg Ala
            340                 345                 350
```

Lys Trp Asn Asn Thr Leu Lys Gln Ile Val Thr Lys Leu Arg Glu Gln
        355   360   365

Phe Gly Asn Asn Lys Thr Thr Ile Val Phe Asn Gln Ser Ser Gly Gly
    370   375   380

Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu Phe Phe
385   390   395   400

Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp His Phe Asn Gly
                405   410   415

Thr Trp Gly Asn Asn Asn Thr Glu Arg Ser Asn Asn Ala Ala Asp Asp
            420   425   430

Asn Asp Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met
        435   440   445

Trp Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro Ile Ser Gly Gln
    450   455   460

Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly
465   470   475   480

Gly Asn Asn Glu Asn Thr Asn Asn Thr Asp Thr Glu Ile Phe Arg Pro
                485   490   495

Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr
            500   505   510

Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys
        515   520   525

Arg Arg Val Val Gln Arg Glu Lys Arg Ala Val Gly Met Leu Gly Ala
    530   535   540

Met Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala
545   550   555   560

Ser Met Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val
                565   570   575

Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu
            580   585   590

Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg Val Leu
        595   600   605

Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly
    610   615   620

Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro Trp Asn Ala Ser
625 630 635 640

Trp Ser Asn Lys Ser Leu Asp Lys Ile Trp Asn Asn Met Thr Trp Met
645 650 655

Glu Trp Glu Arg Glu Ile Asp Asn Tyr Thr Gly Leu Ile Tyr Thr Leu
660 665 670

Ile Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu
675 680 685

Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp Phe Asp Ile Thr Asn
690 695 700

Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met Ile Val Gly Gly Leu Val
705 710 715 720

Gly Leu Arg Ile Val Phe Ala Val Leu Ser Ile Val Asn Arg Val Arg
725 730 735

Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr Arg Leu Pro Ala Pro Arg
740 745 750

Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu Glu Gly Gly Glu Arg Asp
755 760 765

Arg Asp Arg Ser Gly Arg Leu Val Asn Gly Phe Leu Ala Leu Ile Trp
770 775 780

Asp Asp Leu Arg Ser Leu Cys Leu Phe Ser Tyr His Arg Leu Arg Asp
785 790 795 800

Leu Leu Leu Ile Val Ala Arg Ile Val Glu Leu Leu Gly Arg Arg Gly
805 810 815

Trp Glu Ala Leu Lys Tyr Trp Trp Asn Leu Leu Gln Tyr Trp Ser Gln
820 825 830

Glu Leu Lys Asn Ser Ala Val Ser Leu Leu Asn Ala Thr Ala Ile Ala
835 840 845

Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val Val Gln Arg Ala Cys
850 855 860

Arg Ala Ile Leu His Ile Pro Arg Arg Ile Arg Gln Gly Leu Glu Arg
865 870 875 880

Ala Leu Leu

<210> 3

<211> 2562

<212> DNA

<213> Artificial sequence

<220>

<223> Ancestral HIV-1 group M, subtype C, env sequence

<400> 3
atgcgggtga tgggcatcct gcggaactgc cagcagtggt ggatctgggg catcctgggc 60

ttctggatgc tgatgatctg cagcgtgatg ggcaacctgt gggtgaccgt gtactacggc 120

gtgcccgtgt ggaaggaggc caagaccacc ctgttctgcg ccagcgacgc caaggcctac 180

gagcgggagg tgcacaacgt gtgggccacc cacgcctgcg tgcccaccga ccccaacccc 240

caggagatgg tgctggagaa cgtgaccgag aacttcaaca tgtggaagaa cgacatggtg 300

gaccagatgc acgaggacat catcagcctg tgggaccaga gcctgaagcc ctgcgtgaag 360

ctgacccccc tgtgcgtgac cctgaactgc accaacgtga ccaacaccaa caacaacaac 420

aacaccagca tgggcggcga gatcaagaac tgcagcttca acatcaccac cgagctgcgg 480

gacaagaagc agaaggtgta cgccctgttc taccggctgg acatcgtgcc cctgaacgag 540

aacagcaaca gcaacagcag cgagtaccgg ctgatcaact gcaacaccag cgccatcacc 600

caggcctgcc ccaaggtgag cttcgacccc atccccatcc actactgcgc ccccgccggc 660

tacgccatcc tgaagtgcaa caacaagacc ttcaacggca ccggcccctg caacaacgtg 720

agcaccgtgc agtgcaccca cggcatcaag cccgtggtga gcacccagct gctgctgaac 780

ggcagcctgg ccgaggagga gatcatcatc cggagcgaga acctgaccaa caacgccaag 840

accatcatcg tgcacctgaa cgagagcgtg gagatcgtgt gcacccggcc caacaacaac 900

acccggaaga gcatccggat cggccccggc cagaccttct acgccaccgg cgacatcatc 960

ggcgacatcc ggcaggccca ctgcaacatc agcgagaagg agtggaacaa gaccctgcag 1020

cgggtgggca agaagctgaa ggagcacttc cccaacaaga ccatcaagtt cgagcccagc 1080

agcggcggcg acctggagat caccacccac agcttcaact gccggggcga gttcttctac 1140

tgcaacacca gcaagctgtt caacagcacc tacaacagca ccaacaacgg caccaccagc 1200

aacagcacca tcaccctgcc ctgccggatc aagcagatca tcaacatgtg gcagggcgtg 1260

ggccgggcca tgtacgcccc ccccatcgcc ggcaacatca cctgcaagag caacatcacc 1320

ggcctgctgc tgacccggga cggcggcaac accaacaaca ccaccgagac cttccggccc 1380

ggcggcggcg acatgcggga caactggcgg agcgagctgt acaagtacaa ggtggtggag 1440

atcaagcccc tgggcgtggc ccccaccgag gccaagcggc gggtggtgga gcgggagaag 1500

cgggccgtgg gcatcggcgc cgtgttcctg ggcttcctgg gcgccgccgg cagcaccatg 1560

ggcgccgcca gcatcaccct gaccgtgcag gcccggcagc tgctgagcgg catcgtgcag 1620

cagcagagca acctgctgcg ggccatcgag gcccagcagc acatgctgca gctgaccgtg 1680

tggggcatca agcagctgca gacccgggtg ctggccatcg agcggtacct gaaggaccag 1740

cagctgctgg gcatctgggg ctgcagcggc aagctgatct gcaccaccgc cgtgccctgg 1800

aacagcagct ggagcaacaa gagccaggac gacatctggg acaacatgac ctggatgcag 1860

tgggaccggg agatcagcaa ctacaccgac accatctacc ggctgctgga ggacagccag 1920

aaccagcagg agaagaacga gaaggacctg ctggccctgg acagctggaa gaacctgtgg 1980

aactggttcg acatcaccaa ctggctgtgg tacatcaaga tcttcatcat gatcgtgggc 2040

ggcctgatcg gcctgcggat catcttcgcc gtgctgagca tcgtgaaccg ggtgcggcag 2100

ggctacagcc ccctgagctt ccagaccctg acccccaacc cccggggccc cgaccggctg 2160

ggcggcatcg aggaggaggg cggcgagcag gaccgggacc ggagcatccg gctggtgagc 2220

ggcttcctgg ccctggcctg ggacgacctg cggagcctgt gcctgttcag ctaccaccgg 2280

ctgcgggact tcatcctgat cgccgcccgg ggcgtgaacc tgctgggccg gagcagcctg 2340

cggggcctgc agcggggctg ggaggccctg aagtacctgg gcagcctggt gcagtactgg 2400

ggcctggagc tgaagaagag cgccatcagc ctgctggaca ccatcgccat cgccgtggcc 2460

gagggcaccg accggatcat cgagctggtg cagcggatct gccgggccat ccggaacatc 2520

ccccggcgga tccggcaggg cttcgaggcc gccctgcagt ga 2562

<210> 4

<211> 853

<212> PRT

<213> Artificial sequence

<220>

<223> Ancestral HIV-1 group M, subtype C, env sequence.

<400> 4

Met Arg Val Met Gly Ile Leu Arg Asn Cys Gln Gln Trp Trp Ile Trp
1   5   10   15

Gly Ile Leu Gly Phe Trp Met Leu Met Ile Cys Ser Val Met Gly Asn
            20   25   30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Lys
        35   40   45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu Arg Glu Val
    50   55   60

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65 70 75 80

Gln Glu Met Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
85 90 95

Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
100 105 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
115 120 125

Asn Cys Thr Asn Val Thr Asn Thr Asn Asn Asn Asn Asn Thr Ser Met
130 135 140

Gly Gly Glu Ile Lys Asn Cys Ser Phe Asn Ile Thr Thr Glu Leu Arg
145 150 155 160

Asp Lys Lys Gln Lys Val Tyr Ala Leu Phe Tyr Arg Leu Asp Ile Val
165 170 175

Pro Leu Asn Glu Asn Ser Asn Ser Asn Ser Ser Glu Tyr Arg Leu Ile
180 185 190

Asn Cys Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe
195 200 205

Asp Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu
210 215 220

Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val
225 230 235 240

Ser Thr Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln
245 250 255

Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser
260 265 270

Glu Asn Leu Thr Asn Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu
275 280 285

Ser Val Glu Ile Val Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser
290 295 300

Ile Arg Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile
305 310 315 320

Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser Glu Lys Glu Trp Asn
325 330 335

Lys Thr Leu Gln Arg Val Gly Lys Lys Leu Lys Glu His Phe Pro Asn
340 345 350

Lys Thr Ile Lys Phe Glu Pro Ser Ser Gly Gly Asp Leu Glu Ile Thr
355 360 365

Thr His Ser Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser
370 375 380

Lys Leu Phe Asn Ser Thr Tyr Asn Ser Thr Asn Asn Gly Thr Thr Ser
385 390 395 400

Asn Ser Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met
405 410 415

Trp Gln Gly Val Gly Arg Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn
420 425 430

Ile Thr Cys Lys Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly
435 440 445

Gly Asn Thr Asn Asn Thr Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp
450 455 460

Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu
465 470 475 480

Ile Lys Pro Leu Gly Val Ala Pro Thr Glu Ala Lys Arg Arg Val Val
485 490 495

Glu Arg Glu Lys Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe
500 505 510

Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr
515 520 525

Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn
530 535 540

Leu Leu Arg Ala Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val
545 550 555 560

Trp Gly Ile Lys Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg Tyr
565 570 575

Leu Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu
580 585 590

Ile Cys Thr Thr Ala Val Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser
595 600 605

Gln Asp Asp Ile Trp Asp Asn Met Thr Trp Met Gln Trp Asp Arg Glu
    610   615   620

Ile Ser Asn Tyr Thr Asp Thr Ile Tyr Arg Leu Leu Glu Asp Ser Gln
625   630   635   640

Asn Gln Gln Glu Lys Asn Glu Lys Asp Leu Leu Ala Leu Asp Ser Trp
                645   650   655

Lys Asn Leu Trp Asn Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile
            660   665   670

Lys Ile Phe Ile Met Ile Val Gly Gly Leu Ile Gly Leu Arg Ile Ile
        675   680   685

Phe Ala Val Leu Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro
    690   695   700

Leu Ser Phe Gln Thr Leu Thr Pro Asn Pro Arg Gly Pro Asp Arg Leu
705   710   715   720

Gly Gly Ile Glu Glu Glu Gly Gly Glu Gln Asp Arg Asp Arg Ser Ile
                725   730   735

Arg Leu Val Ser Gly Phe Leu Ala Leu Ala Trp Asp Asp Leu Arg Ser
            740   745   750

Leu Cys Leu Phe Ser Tyr His Arg Leu Arg Asp Phe Ile Leu Ile Ala
        755   760   765

Ala Arg Gly Val Asn Leu Leu Gly Arg Ser Ser Leu Arg Gly Leu Gln
    770   775   780

Arg Gly Trp Glu Ala Leu Lys Tyr Leu Gly Ser Leu Val Gln Tyr Trp
785   790   795   800

Gly Leu Glu Leu Lys Lys Ser Ala Ile Ser Leu Leu Asp Thr Ile Ala
                805   810   815

Ile Ala Val Ala Glu Gly Thr Asp Arg Ile Ile Glu Leu Val Gln Arg
            820   825   830

Ile Cys Arg Ala Ile Arg Asn Ile Pro Arg Arg Ile Arg Gln Gly Phe
        835   840   845

Glu Ala Ala Leu Gln
    850

<210> 5

<211> 2652

<212> DNA

<213> Artificial sequence

<220>

<223> Semi-optimized ancestral viral sequences for HIV-1 subtypes B and C

<400> 5
atgagagtga aggggatcag gaagaactat cagcacttgt ggagatgggg caccatgctc 60
cttgggatgt tgatgatctg tagcgccgcc gagaagctgt gggtgaccgt gtactacggc 120
gtgcccgtgt ggaaggaggc caccaccacc ctgttctgcg ccagcgacgc caaggcttac 180
gacaccgagg tccacaacgt gtgggccacc cacgcctgcg tgcccaccga ccccaacccc 240
caggaggtgg tgctggagaa cgtgaccgag aacttcaaca tgtggaagaa caacatggtg 300
gagcagatgc acgaggacat catcagcctg tgggaccaga gcctgaagcc ctgcgtgaag 360
ttaacccccc tgtgcgtgac cctgaactgc accgacgacc tgcgcaccaa cgccaccaac 420
accaccaaca gcagcgccac caccaacacc accagcagcg gcggcggcac gatggagggc 480
gagaagggcg agatcaagaa ctgcagcttc aacgtgacca ccagcatccg cgacaagatg 540
cagaaggagt acgccctgtt ctacaagctg gacgtggtgc ccatcgacaa cgacaacaac 600
aacaccaaca acaacaccag ctaccgcctc atcaactgca acaccagcgt gatcacccag 660
gcctgcccca aggtgagctt cgagcccatc cccatccact actgcacccc cgccggcttc 720
gccatcctga agtgcaacga caagaagttc aacggcaccg gcccctgcac caacgtgagc 780
accgtgcagt gcacccacgg catccgcccc gtggtgagca cccagctgct gctgaacggc 840
agcctggccg aggaggaggt ggtgatccgc agcgagaact tcaccgacaa cgccaagacc 900
atcatcgtgc agctgaacga gagcgtggag atcaactgca cgcgtcccaa caacaacacc 960
cgcaagagca tccccatcgg ccctggccgc gccctgtacg ccaccggcaa gatcatcggc 1020
gacatccgcc aggcccactg caacctgtcg cgagccaagt ggaacaacac cctgaagcag 1080
atcgtgacca agctgcgcga gcagttcggc aacaacaaga ccaccatcgt gttcaaccag 1140
agcagcggcg gcgaccccga gatcgtgatg cacagcttca actgcggcgg cgaattcttc 1200
tactgcaaca gcacccagct gttcaacagc acctggcact tcaacggcac ctggggcaac 1260
aacaacaccg agcgcagcaa caacgccgcc gacgacaacg acaccatcac cctgccctgc 1320
cgcatcaagc agatcatcaa catgtggcag gaggtgggca aggccatgta cgcccccccc 1380
atcagcggcc agatccgctg cagcagcaac atcaccggcc tgctgctgac tcgagacggc 1440
ggcaacaacg agaacaccaa caacaccgac accgagatct tccgccccgg gggcggcgac 1500
atgcgcgaca actggcgcag cgagctgtac aagtacaagg tggtgaagat cgagcccctg 1560
ggcgtagcac ccaccaaggc aaagagaaga gtggtgcaga gagaaaaaag cgcagtggga 1620
atgctaggag ctatgttcct tgggttcttg ggagcagcag gaagcactat gggcgcagcg 1680

tcaatgacgc tgaccgtaca ggccagacaa ttattgtctg gtatagtgca gcagcagaac 1740

aatctgctga gggctattga ggcgcaacag catctgttgc aactcacagt ctggggcatc 1800

aagcagctcc aggcaagagt cctggctgtg gaaagatacc taaaggatca gcagctcctg 1860

gggatttggg gttgctctgg aaaactcatc tgcaccactg ctgtgccttg gaatgctagc 1920

tggagcaaca agagcctgga caagatctgg aacaacatga cctggatgga gtgggagcgc 1980

gagatcgaca actacaccgg cctgatctac accctgatcg aggagagcca gaaccagcag 2040

gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctgtg gaactggttc 2100

gatatcacca actggctgtg gtacatcaag atcttcatca tgatcgtggg cggcctggtg 2160

ggcctgcgca tcgtgttcgc cgtgctgagc atcgtgaacc gcgtgcgcca gggctacagc 2220

cccctgagct tccagaccca cctgccagcc ccgaggggac ccgacaggcc cgaaggaatc 2280

gaagaagaag gtggagagag agacagagac agatccggtc gattagtgaa tggattctta 2340

gcacttatct gggacgacct gcggagcctg tgcctcttca gctaccaccg cttgagcgac 2400

ttactcttga ttgtagcgag gattgtggaa cttctgggac gcagggggtg ggaggccctc 2460

aaatattggt ggaatctcct gcagtactgg agtcaggaac taaagaatag cgccgtgagc 2520

ctgctgaacg ccaccgccat cgccgtggcc gagggcaccg accgcgtgat cgaggtggtg 2580

cagcgcgcct gccgcgccat cctgcacatc ccccgccgca tccgccaggg cctggagcgc 2640

gccctgctgt ga 2652

<210> 6

<211> 2562

<212> DNA

<213> Artificial sequence

<220>

<223> Semi-optimized ancestral viral sequences for HIV-1 subtypes B and C

<400> 6
atgagagtga tggggatact gaggaattgt caacaatggt ggatatgggg catcctaggc 60

ttttggatgc taatgatttg tgacgtgatg ggcaacctgt gggtgaccgt gtactacggc 120

gtgcccgtgt ggaaggaggc caagaccacc ctgttctgcg ccagcgacgc caaggcctac 180

gagcgggagg tgcacaacgt gtgggccacc cacgcctgcg tgcccaccga ccccaacccc 240

caggagatgg tgctggagaa cgtgaccgag aacttcaaca tgtggaagaa cgacatggtg 300

gaccagatgc acgaggacat catcagcctg tgggaccaga gcctgaagcc ctgcgtgaag 360

ctgacccccc tgtgcgtgac cctgaactgc accaacgtga ccaacaccaa caacaacaac 420

aacaccagca tgggcggcga gatcaagaac tgcagcttca acatcaccac cgagctgcgg 480

gacaagaagc agaaggtgta cgccctgttc taccggctgg acatcgtgcc cctgaacgag 540

aacagcaaca gcaacagcag cgagtaccgg ctgatcaact gcaacaccag cgccatcacc 600
caggcctgcc ccaaggtgag cttcgacccc atccccatcc actactgcgc ccccgccggc 660
tacgccatcc tgaagtgcaa caacaagacc ttcaacggca ccggcccctg caacaacgtg 720
agcaccgtgc agtgcaccca cggcatcaag cccgtggtga gcacccagct gctgctgaac 780
ggcagcctgg ccgaggagga gatcatcatc cggagcgaga acctgaccaa caacgccaag 840
accatcatcg tgcacctgaa cgagagcgtg gagatcgtgt gcacccggcc caacaacaac 900
acccggaaga gcatccggat cggccccggc cagaccttct acgccaccgg cgacatcatc 960
ggcgacatcc ggcaggccca ctgcaacatc agcgagaagg agtggaacaa gaccctgcag 1020
cgggtgggca agaagctgaa ggagcacttc cccaacaaga ccatcaagtt cgagcccagc 1080
agcggcggcg acctggagat caccacccac agcttcaact gccggggcga gttcttctac 1140
tgcaacacca gcaagctgtt caacagcacc tacaacagca ccaacaacgg caccaccagc 1200
aacagcacca tcaccctgcc ctgccggatc aagcagatca tcaacatgtg gcagggcgtg 1260
ggccgggcca tgtacgcccc ccccatcgcc ggcaacatca cctgcaagag caacatcacc 1320
ggcctgctgc tgacccggga cggcggcaac accaacaaca ccaccgagac cttccggccc 1380
ggcggcggcg acatgcggga caactggcgg agcgagctgt acaagtacaa ggtggtggag 1440
atcaagcccc tgggcgtagc acccactgag gcaaaaagga gagtggtgga gagagaaaaa 1500
agagcagtgg gaataggagc tgtgttcctt gggttcttgg gagcagcagg aagcactatg 1560
ggcgcggcgt caataacgct gacggtacag gccagacaat tattgtctgg tatagtgcaa 1620
cagcaaagca atttgctgag ggctatagag gcgcaacagc atatgttgca actcacggtc 1680
tggggcatta agcagctcca gacaagagtc ctggctatag aaagatacct aaaggatcag 1740
cagctcctgg gcatttgggg ctgctctgga aaactcatct gcaccactgc tgtgccttgg 1800
aactctagct ggagcaacaa gagccaggac gacatctggg acaacatgac ctggatgcag 1860
tgggaccggg agatcagcaa ctacaccgac accatctacc ggctgctgga ggacagccag 1920
aaccagcagg agaagaacga gaaggacctg ctggccctgg acagctggaa gaacctgtgg 1980
aactggttcg acatcaccaa ctggctgtgg tacatcaaga tcttcatcat gatcgtgggc 2040
ggcctgatcg gcctgcggat catcttcgcc gtgctgagca tcgtgaaccg ggtgcggcag 2100
ggctacagcc ccctgagctt ccagaccctt accccaaacc cgaggggacc cgacaggctc 2160
ggaggaatcg aagaagaagg tggagagcaa gacagagaca gatccattcg attagtgagc 2220
ggattcttag cactggcctg ggacgacctg cggagcctgt gcctcttcag ctaccaccga 2280
ttgagagact tcatattgat tgcagccaga gggtgggaac ttctgggacg cagcagtctc 2340
aggggactgc agaggggggtg ggaagccctt aagtatctgg gaagtcttgt gcagtattgg 2400
ggtctggagc taaaaaagag tgctattagc ctgctggaca ccatcgccat cgccgtggcc 2460
gagggcaccg accggatcat cgagctggtg cagcggatct gccgggccat ccggaacatc 2520
ccccggcgga tccggcaggg cttcgaggcc gccctgcagt ga 2562

<210> 7

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence-maximum likelihood reconstruction of determine
d ancestral node.

<400> 7
gatcctg 7

<210> 8

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, most parsimonious reconstruction of determine
d ancestral node

<220>

<221> variation

<222> (3)..(3)

<223> W can be an A or T

<400> 8
gawcctg 7

<210> 9

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 9
gaacctg 7

<210> 10

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 10
gaaactc   7

<210> 11

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 11
gatactc   7

<210> 12

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, most parsimonious reconstruction of determine
d ancestral node.

<220>

<221> variation

<222> (3)..(3)

<223> W can be an A or T

<400> 12
gawactc   7

<210> 13

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 13
catactc 7

<210> 14

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 14
catactt 7

<210> 15

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 15
catacta 7

<210> 16

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 16
catattg 7

<210> 17

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, most parsimonious reconstruction of determine
d ancestral node.

<220>

<221> variation

<222> (7)..(7)

<223> V can also be an A, C or G

<400> 17
catactv 7

<210> 18

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 18
catgctg 7

<210> 19

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 19
catactg   7

<210> 20

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 20
caagctg   7

<210> 21

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 21
catgctg   7

<210> 22

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 22
cttgctg   7

<210> 23

<211> 7

<212> DNA

<213> Artificial sequence

<220>

<223> Consensus sequence, maximum likelihood reconstruction of determin
ed ancestral node.

<400> 23
cttgctt 7

<210> 24

<211> 1503

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor of reconstruction clade B gag gene se
quence

<400> 24
atgggtgcga gagcgtcagt attaagcggg ggagaattag ataaatggga aaaaattcgg 60

ttacggccag ggggaaagaa aaaatataaa ttaaaacata tagtatgggc aagcagggag 120

ctagaacgat tcgcagttaa tcctggcctt ttagaaacat cagaaggctg tagacaaata 180

ctgggacagc tacaaccatc ccttcagaca ggatcagaag aacttagatc attatataat 240

acagtagcag tcctctattg tgtgcatcaa aagatagagg taaaagacac caaggaagct 300

ttagataaga tagaggaaga gcaaaacaaa agtaagaaaa aggcacagca agcagcagct 360

gacacaggaa acagcagcca ggtcagccaa aattacccta tagtgcagaa cctacagggg 420

caaatggtac atcaggccct atcacctaga actttaaatg catgggtaaa agtaatagaa 480

gagaaggctt tcagcccaga agtaataccc atgttttcag cattatcaga aggagccacc 540

ccacaagatt taaacaccat gctaaacaca gtggggggac atcaagcagc catgcaaatg 600

ttaaaagaga ccatcaatga ggaagctgca gaatgggata gattgcatcc agtgcatgca 660

gggcctattg caccaggcca gatgagagaa ccaaggggaa gtgacatagc aggaactact 720

agtacccttc aggaacaaat agcatggatg acaaataatc cacctatccc agtaggagaa 780

atctataaaa gatggataat cctgggatta aataaaatag taagaatgta tagccctgtc 840

agcattctgg acataagaca aggaccaaag gaacccttta gagactatgt agaccggttc 900

tataaaactc taagagccga gcaagcttca caggaggtaa aaaattggat gacagaaacc 960

ttgttggtcc aaaatgcgaa cccagattgt aagactatct taaaagcatt gggaccagga 1020

gctacactag aagaaatgat gacagcatgt cagggagtgg ggggacccgg ccataaagca 1080

agagttttgg ctgaagcaat gagccaagta acaaattcag ctaccataat gatgcagaga 1140

ggcaatttta ggaacccaag aaagactgtt aagtgtttca attgtggcaa agaagggcac 1200

atagccagaa attgcagggc ccctaggaaa aagggctgtt ggaaatgtgg aaaggaagga 1260

caccaaatga aagattgtac tgagagacag gctaattttt tagggaaaat ctggccttcc 1320

cacaagggaa ggccagggaa ttttcttcag agcagaccag agccaacagc cccaccagaa 1380

gagagcttca ggtttgggga agagacaaca actccctctc agaagcagga gcagaaagac 1440

aaggaactgt atcctttagc ttccctcaaa tcactctttg gcaacgaccc ctcgtcacaa 1500

taa 1503

<210> 25

<211> 1503

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade B gag gene s equence

<400> 25

atgggtgcga gagcgtcagt attaagcggg ggagaattag atagatggga aaaaattcgg 60

ttaaggccag ggggaaagaa aaaatataga ttaaaacata tagtatgggc aagcagggag 120

ctagaacgat tcgcagttaa tcctggcctg ttagaaacat cagaaggctg tagacaaata 180

ctgggacagc tacaaccatc ccttcagaca ggatcagaag aacttagatc attatataat 240

acagtagcaa ccctctattg tgtgcatcaa aggatagagg taaaagacac caaggaagct 300

ttagagaaga tagaggaaga gcaaaacaaa agtaagaaaa aggcacagca agcagcagct 360

gacacaggaa acagcagcca ggtcagccaa aattacccta tagtgcagaa cctccagggg 420

caaatggtac atcaggccat atcacctaga actttaaatg catgggtaaa agtagtagag 480

gagaaggctt tcagcccaga agtaataccc atgttttcag cattatcaga aggagccacc 540

ccacaagatt taaacaccat gctaaacaca gtggggggac atcaagcagc catgcaaatg 600

ttaaaagaga ccatcaatga ggaagctgca gaatgggata gattgcatcc agtgcatgca 660

gggcctattg caccaggcca gatgagagaa ccaaggggaa gtgacatagc aggaactact 720

agtacccttc aggaacaaat aggatggatg acaaataatc cacctatccc agtaggagaa 780

atctataaaa gatggataat cctgggatta aataaaatag taagaatgta tagccctacc 840

agcattctgg acataagaca aggaccaaag gaacccttta gagactatgt agaccggttc 900

tataaaactc taagagccga gcaagcttca caggaggtaa aaaattggat gacagaaacc 960

ttgttggtcc aaaatgcgaa cccagattgt aagactattt taaaagcatt gggaccagca 1020

gctacactag aagaaatgat gacagcatgt cagggagtgg ggggacccgg ccataaagca 1080

agagttttgg ctgaagcaat gagccaagta acaaattcag ctaccataat gatgcagaga 1140

ggcaatttta ggaaccaaag aaagactgtt aagtgtttca attgtggcaa agaagggcac 1200

atagccaaaa attgcagggc ccctaggaaa aagggctgtt ggaaatgtgg aaaggaagga 1260

caccaaatga aagattgtac tgagagacag gctaattttt tagggaagat ctggccttcc 1320

cacaagggaa ggccagggaa ttttcttcag agcagaccag agccaacagc cccaccagaa 1380

gagagcttca ggtttgggga agagacaaca actccctctc agaagcagga gccgatagac 1440

aaggaactgt atcctttagc ttccctcaga tcactctttg gcaacgaccc ctcgtcacaa 1500

taa 1503

<210> 26

<211> 1503

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade B gag gen e sequence

<400> 26
atgggtgcgg gagcgtcggt attaagcggg ggaaaattag ataggtggga aaaaattcgg 60

ttaaggccag ggggaaagaa aaaatataaa ttaaaacata tagtatgggc aagcagggag 120

ctagaacgat ttgcagtcaa tcctggcctg ttagaaacat cagaaggctg cagacgaata 180

ctggaacagc tacatccatc ccttcagaca ggatcagaag aacttaaatc attatataat 240

acggtagcaa ccctctattg tgtgcatcaa aatatagagg taagagacac caaggatgct 300

ttagaaaaaa tagaggaaga acaaaacaaa attaagaaaa gggcacagca agcagcagct 360

gacacaggaa acagcaaccc ggtcagccaa aattacccta tagtgcagaa tatgcagggg 420

caaatggtac atcaggccat atcacctaga actttaaatg catgggtaaa agtagtagaa 480

gagaaggctt tcagccccga agtaataccc atgttttcag cattatcaga aggagccacc 540

ccacaagatt taaacaccat gctaaacaca gtggggggac atcaagcagc catgcaaatg 600

ttaaaagaaa ccatcaatga ggaagctgca gaatgggata gattgcaccc agtgcatgca 660

gggcctattg caccaggcca gatgagagaa ccaaggggaa gtgacatagc aggaactact 720

agtacccttc aggaacaaat aggatggatg acacataatc cacctatccc agtaggagaa 780

atctataaaa gatggataat catgggatta aataaaatag taagaatgta tagccctacc 840

agcattctgg acataagaca aggaccaaag gaacccttta gagattatgt tgaccggttc 900

tataaaactc taagagccga gcaagcttca caggaggtaa aaaattggat gacagaaacc 960

ttgttggtcc aaaatgcgaa cccagattgt aagaccattt taaaagcatt aggaccagca 1020

gctacactag aagaaatgat gacagcatgt cagggagtgg gagggcccag ccataaagca 1080

agagttttgg cagaagcaat gagccaagca acaaattcag ctaccataat gatgcagagg 1140

ggcaatttta agggccaaag aaagactgtt aaatgtttca attgtggcaa agaagggcac 1200

atagccagaa attgcagggc ccctagaaaa aagggctgtt ggaaatgtgg aaaggaagga 1260

caccaaatga aagattgtac tgagagacag gctaattttt tagggaagat ctggccttcc 1320

cacaagggaa ggccagggaa ttttctccaa agcaggccag agccaacagc cccaccagaa 1380

gagagcttca ggtttgggga ggagacaaca actccccctc agaagcagga gccgagggac 1440

aaggaacagt atcccttgac ttccctcaga tcactctttg gcaacgaccc atcgtcacaa 1500

taa 1503

<210> 27

<211> 2589

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B env gene se
quence

<400> 27

atgagagtga aggggatcag gaagaattgt cagcacttgt ggaaatgggg caccatgctc 60

cttgggatgt tgatgatctg tagtgctgca gaaaacttgt gggtcacagt ctattatggg 120

gtacctgtgt ggaaagaagc aaccaccact ctattttgtg catcagatgc taaagcatat 180

aaaacagagg tacataatgt ctgggccaca catgcctgtg tacccacaga ccccaaccca 240

caagaagtag tattggaaaa tgtgacagaa aattttaaca tgtggaaaaa taacatggta 300

gaacagatgc atgaggatat aatcagttta tgggatcaaa gcctaaagcc atgtgtaaaa 360

ttaaccccac tctgtgttac tttaaattgc actgatgcga acaagaatgc tactaatacc 420

aatagtagta gtgggggaac aatggagaaa ggagaaatga aaaactgctc tttcaatatc 480

accacaagca taagagataa gatgcagaaa gaatatgcac ttttttataa acttgatgta 540

gtaccaatag ataatgataa taatagtaat aataatacca actataggtt gataaattgt 600

aatacctcag tcattacaca ggcctgtcca aaggtatcct ttgagccaat tcccatacat 660

tattgtaccc cggctggttt tgcgattcta aagtgtaatg ataagaagtt caatggaaca 720

ggaccatgta aaaatgtcag cacagtacaa tgtacacatg gaattaggcc agtagtgtca 780

actcaactgc tgttaaatgg cagtctagca gaagaagagg tagtaattag atctgaaaat 840

ttcacggaca atgctaaaac cataatagta cagctgaatg aatctgtaga aattaattgt 900

acaagaccca acaacaatac aagaaaaagt atacctatag gaccagggag agcactttat 960

acaacaggag aaataatagg agatataaga caagcacatt gtaacattag tagagcaaaa 1020

tggaataaca ctttaaaaca ggtagttaca aaattaagag aacaatttgg gaataataaa 1080

acaatagtct ttaatccatc ctcaggaggg gacccagaaa ttgtaatgca cagttttaat 1140

tgtggagggg aatttttcta ctgtaataca acacaactgt ttaatagtac ttggaatagt 1200

actgaagggt caaataaaac tacagggtca aataacactg gaggagaaac tatcacactc 1260

ccatgcagaa taaaacaaat tataaacatg tggcaggaag taggaaaagc aatgtatgcc 1320

cctcccatca gaggacaaat taaatgttca tcaaatatta cagggctact attaacaaga 1380

gatggtggtg aaaatagtac caatgagacc gagatcttca gacctggagg aggagatatg 1440

agggacaatt ggagaagtga attatataaa tataaagtag taaaaattga accattagga 1500

gtagcaccca ccaaggcaaa gagaagagtg gtgcaaagag aaaaaagagc agtgggaata 1560

ataggagcta tgttccttgg gttcttggga gcagcaggaa gcactatggg cgcagcgtca 1620

atgacgctga cggtacaggc cagacaatta ttgtctggta tagtgcaaca gcaaaacaat 1680

ttgctgaggg ctattgaggc gcaacagcat ctgttgcaac tcacggtctg gggcatcaaa 1740

cagctccagg caagagtcct ggctgtggaa agatacctaa gggatcaaca gctcctagga 1800

atttggggtt gctctggaaa actcatttgc accactactg tgccttggaa tgctagttgg 1860

agtaataaat ctctggataa gatttggaat aacatgacct ggatggagtg ggaaagagaa 1920

attgacaatt acacaggctt aatatacaac ttaattgaag aatcgcagaa ccagcaagaa 1980

aagaatgaac aagaattatt ggaattggat aagtgggcaa gtttgtggaa ttggtttgac 2040

ataacacaat ggctgtggta tataaaaata ttcataatga tagtaggagg cttggtaggt 2100

ttaagaatag tttttgctgt gctttctata gtgaatagag ttaggcaggg atactcacca 2160

ttatcatttc agacccgcct cccagccccg aggggacccg acaggcccga aggaatcgaa 2220

gaagaaggtg gagagagaga cagagacaga tccggtcgat tagtgaatgg attcttagca 2280

cttatctggg acgatctgcg gagcctgtgc ctcttcagct accaccgctt gagagactta 2340

ctcttgattg tagcgaggat tgtggaactt ctgggacgca gggggtggga agccctcaaa 2400

tattggtgga atctcctgca gtattggagt caggaactaa agaatagtgc tgttagcttg 2460

cttaatgcca cagcaatagc agtagctgag gggacagata gggttataga agtagtacaa 2520

agagcttgta gagctattct tcacatacct agaagaataa gacagggctt agaaagggct 2580

ttgctataa 2589

<210> 28

<211> 2589

<212> DNA

<213> Artificial sequence

<220>

<223> Least square center of tree and minimum of means center of tree r

econstruction of clade B env gene sequence

<400> 28
atgagagtga aggggatcag gaagaattat cagcacttgt ggagatgggg caccatgctc 60

cttgggatgt tgatgatctg tagtgctgca gaaaaattgt gggtcacagt ctattatggg 120

gtacctgtgt ggaaagaagc aaccaccact ctattttgtg catcagatgc taaagcatat 180

gatacagagg tacataatgt ttgggccaca catgcctgtg tacccacaga ccccaaccca 240

caagaagtag tattggaaaa tgtgacagaa aattttaaca tgtggaaaaa taacatggta 300

gaacagatgc atgaggatat aatcagttta tgggatcaaa gcctaaagcc atgtgtaaaa 360

ttaaccccac tctgtgttac tttaaattgc actgatttga ataagaatgc tactaatacc 420

aatagtagta gcggggaaat gatggagaaa ggagaaataa aaaactgctc tttcaatatc 480

accacaagca taagagataa ggtgcagaaa gaatatgcac tttttataa acttgatgta 540

gtaccaatag ataatgataa taatactaat aatactacca gctataggtt gataagttgt 600

aacacctcag tcattacaca ggcctgtcca aaggtatcct ttgagccaat tcccatacat 660

tattgtgccc cggctggttt tgcgattcta aagtgtaatg ataagaagtt caatggaaca 720

ggaccatgta caaatgtcag cacagtacaa tgtacacatg gaattaggcc agtagtatca 780

actcaactgc tgttaaatgg cagtctagca gaagaagagg tagtaattag atctgacaat 840

ttcacggaca atgctaaaac cataatagta cagctgaatg aatctgtaga aattaattgt 900

acaagaccca acaacaatac aagaaaaagt atacatatag gaccagggag agcattttat 960

acaacaggag aaataatagg agatataaga caagcacatt gtaacattag tagagcaaaa 1020

tggaataaca ctttaaaaca gatagttaaa aaattaagag aacaatttgg gaataataaa 1080

acaatagtct ttaatcaatc ctcaggaggg gacccagaaa ttgtaatgca cagttttaat 1140

tgtggagggg aatttttcta ctgtaattca acacaactgt ttaatagtac ttggaatggt 1200

acttggactt ggaatactac tgaagggtca aatgacactg aaggagacac tatcacactc 1260

ccatgcagaa taaaacaaat tataaacatg tggcaggaag taggaaaagc aatgtatgcc 1320

cctcccatca gaggacaaat tagatgttca tcaaatatta cagggctgct attaacaaga 1380

gatggtggta ataataacac caacgagacc gagatcttca gacctggagg aggagatatg 1440

agggacaatt ggagaagtga attatataaa tataaagtag taaaaattga accattagga 1500

gtagcaccca ccaaggcaaa gagaagagtg gtgcagagag aaaaaagagc agtgggaata 1560

ataggagctg tgttccttgg gttcttggga gcagcaggaa gcactatggg cgcagcgtca 1620

atgacgctga cggtacaggc cagacaatta ttgtctggta tagtgcaaca gcagaacaat 1680

ttgctgaggg ctattgaggc gcaacagcat ctgttgcaac tcacagtctg gggcatcaag 1740

cagctccagg caagagtcct ggctgtggaa agatacctaa gggatcaaca gctcctgggg 1800

atttggggtt gctctggaaa actcatttgc accactgctg tgccttggaa tgctagttgg 1860

agtaataaat ctctggatga gatttggaat aacatgacct ggatggagtg ggaaagagaa 1920

attgacaatt acacaagctt aatatacacc ttaattgaag aatcgcaaaa ccaacaagaa 1980

aagaatgaac aagaattatt ggaattagat aaatgggcaa gtttgtggaa ttggtttgac 2040

ataacaaact ggctgtggta tataaaaata ttcataatga tagtaggagg cttggtaggt 2100

ttaagaatag tttttgctgt actttctata gtgaatagag ttaggcaggg atactcacca 2160

ttatcgtttc agacccgcct cccagccccg aggggacccg acaggcccga aggaatcgaa 2220

gaagaaggtg gagagagaga cagagacaga tccggtcgat tagtgaacgg attcttagca 2280

cttatctggg acgacctgcg gagcctgtgc ctcttcagct accaccgctt gagagactta 2340

ctcttgattg taacgaggat tgtggaactt ctgggacgca gggggtggga agccctcaaa 2400

tattggtgga atctcctaca gtattggagt caggaactaa agaatagtgc tgttagcttg 2460

ctcaatgcca cagccatagc agtagctgag gggacagata gggttataga agtagtacaa 2520

agagcttgta gagctattct ccacatacct acaagaataa gacagggctt ggaaagggct 2580

ttgctataa 2589

<210> 29

<211> 621

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B nef gene se
quence

<400> 29
atgggtggca agtggtcaaa acgtagtgtg gttggatggc ctgctgtaag ggaaagaatg 60

agacgagctg agccagcagc agatggggtg ggagcagtat ctcgagacct ggaaaaacat 120

ggagcaatca caagtagcaa tacagcagct actaatgctg cttgtgcctg gctagaagca 180

caagaggagg aggaggtggg ttttccagtc agacctcagg tacctttaag accaatgact 240

tacaaggcag ctgtagatct tagccacttt ttaaaagaaa aggggggact ggaagggcta 300

gtttactccc aaaaaagaca agatatcctt gatctgtggg tctaccacac acaaggctac 360

ttccctgatt ggcagaacta cacaccaggg ccagggacca gatatccact gacctttgga 420

tggtgcttca agctagtacc agttgagcca gagaaggtag aagaggccac tgaaggagag 480

aacaacagct tgttacaccc tatgagcctg catggaatgg atgacccgga gagagaagtg 540

ttagtgtgga ggtttgacag ccgcctagca tttcatcaca tggcccgaga gaagcatccg 600

gagtactaca aggactgctg a 621

<210> 30

<211> 621

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade B nef gene s
equence

<400> 30
atgggtggca agtggtcaaa acgtagtgtg gttggatggc ctgctgtaag ggaaagaatg 60

agacgagctg agccagcagc agatggggtg ggagcagtat ctcgagacct ggaaaaacat 120

ggagcaatca caagtagcaa tacagcagct actaatgctg attgtgcctg gctagaagca 180

caagaggagg aggaggtggg ttttccagtc agacctcagg tacctttaag accaatgact 240

tacaaggcag ctttagatct tagccacttt ttaaaagaaa aggggggact ggaagggcta 300

atttactccc aaaaaagaca agatatcctt gatctgtggg tctaccacac acaaggctac 360

ttccctgatt ggcagaacta cacaccaggg ccagggatca gatatccact gacctttgga 420

tggtgcttca agctagtacc agttgagcca gagaaggtag aagaggccaa tgaaggagag 480

aacaacagct tgttacaccc tatgagcctg catgggatgg atgacccgga gaaagaagtg 540

ttagtgtgga agtttgacag ccgcctagca tttcatcaca tggcccgaga gctgcatccg 600

gagtactaca aggactgctg a 621

<210> 31

<211> 621

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade B nef gen
e sequence

<400> 31
atgggtggca agtggtcaaa acgtagtgtg gttggatggc ctgctgtaag ggaaagaatg 60

agacgagctg agccagcagc agatggggtg ggagcagtat ctcgagacct ggaaaaacat 120

ggagcaatca caagtagcaa tacagcagct actaatgctg attgtgcctg gctagaagca 180

caagaggagg aggaggtggg ttttccagtc agacctcagg tacctttaag accaatgact 240

tacaaggcag ctttagatct tagccacttt ttaaaagaaa aggggggact ggaagggcta 300

atttactccc aaaaaagaca agatatcctt gatctgtggg tctaccacac acaaggctac 360

ttccctgatt ggcagaacta cacaccaggg ccagggatca gatatccact gacctttgga 420

tggtgcttca agctagtacc agttgagcca gagaaggtag aagaggccaa tgaaggagag 480

aacaactgct tgttacaccc tatgagccag catgggatgg atgacccgga gaaagaagtg 540

ttagtgtgga agtttgacag ccgcctagca tttcatcaca tggcccgaga gctgcatccg 600

gagtactaca aggactgctg a 621

<210> 32

<211> 3012

<212> DNA

<213> Artificial Sequence

<220>

<223> Most recent common ancestor reconstruction of clade B pol gene se
quence

<400> 32
tttttaggg aaaatctggc cttcccacaa gggaaggcca gggaactttc ttcagagcag 60

accagagcca acagccccac cagaagagag cttcaggttt ggggaagaga caacaactcc 120

ctctcagaag caggagcaga tagacaagga actgtatcct ttagcttccc tcaaatcact 180

ctttggcaac gacccctcgt cacaataaag ataggggggc aactaaagga agctctatta 240

gatacaggag cagatgatac agtattagaa gaaatgaatt tgccaggaaa atggaaacca 300

aaaatgatag ggggaattgg aggttttatc aaagtaagac agtatgatca aatacccata 360

gaaatctgtg gacataaagc tataggtaca gtattagtag gacctacacc tgtcaacata 420

attggaagaa atctgttgac tcagattggt tgcactttaa attttcccat tagtcctatt 480

gaaactgtac cagtaaaatt aaagccagga atggatggcc caaaagttaa acaatggcca 540

ttgacagaag aaaaaataaa agcattagta gaaatttgta cagaaatgga aaaggaagga 600

aaaatttcaa aaattgggcc tgaaaatcca tacaatactc cagtatttgc cataaagaaa 660

aaagacagta ctaaatggag aaaattagta gatttcagag aacttaataa gagaactcaa 720

gacttctggg aagttcaatt aggaatacca catcctgcag ggttaaaaaa gaaaaaatca 780

gtaacagtac tggatgtggg tgatgcatat ttttcagttc ccttagatga agacttcagg 840

aagtatactg catttaccat acctagtata aacaatgaga caccagggat tagatatcag 900

tacaatgtgc ttccacaggg atggaaagga tcaccagcaa tattccaaag tagcatgaca 960

aaaatcttag agccttttag aaaacaaaat ccagaaatag ttatctatca atacatggat 1020

gatttgtatg taggatctga cttagaaata gggcagcata gaacaaaaat agaggaactg 1080

agagaacatc tgttgaggtg gggatttacc acaccagaca aaaaacatca gaaagaacct 1140

ccatttcttt ggatgggtta tgaactccat cctgataaat ggacagtaca gcctatagtg 1200

ctgccagaaa aagacagctg gactgtcaat gacatacaga agttagtggg aaaattgaat 1260

tgggcaagtc agatttatgc agggattaaa gtaaagcaat tatgtaaact ccttagggga 1320

accaaagcac taacagaagt agtaccacta acagaagaag cagagctaga actggcagaa 1380

aacagggaga ttctaaaaga accagtacat ggagtgtatt atgacccatc aaaagactta 1440

atagcagaaa tacagaagca ggggcaaggc caatggacat atcaaattta tcaagagcca 1500
tttaaaaatc tgaaaacagg aaagtatgca agaatgaggg gtgcccacac taatgatgta 1560
aaacaattaa cagaggcagt gcaaaaaata gccacagaaa gcatagtaat atggggaaag 1620
actcctaaat ttaaactacc catacaaaag gaaacatggg aagcatggtg gacagagtat 1680
tggcaagcca cctggattcc tgagtgggag tttgtcaata cccctccctt agtaaaatta 1740
tggtaccagt tagagaaaga acccatagta ggagcagaaa ctttctatgt agatggggca 1800
gctaatagag agactaaatt aggaaaagca ggatatgtta ctgacagagg aagacaaaaa 1860
gttgtctccc taactgacac aacaaatcag aagactgagt tacaagcaat tcatctagct 1920
ttgcaggatt cgggattaga agtaaacata gtaacagact cacaatatgc attaggaatc 1980
attcaagcac aaccagataa gagtgaatca gagttagtca gtcaaataat agagcagtta 2040
ataaaaaagg aaaaggtcta cctggcatgg gtaccagcac acaaaggaat tggaggaaat 2100
gaacaagtag ataaattagt cagtactgga atcaggaaag tactattttt ggatggaata 2160
gataaggccc aagaagaaca tgagaaatat cacagtaatt ggagagcaat ggctagtgat 2220
tttaacctgc cacctgtagt agcaaaagaa atagtagcca gctgtgataa atgtcagcta 2280
aaaggagaag ccatgcatgg acaagtagac tgtagtccag gaatatggca actagattgt 2340
acacatttag aaggaaaagt tatcctggta gcagttcatg tagccagtgg ctatatagaa 2400
gcagaagtta ttccagcaga aacagggcag gaaacagcat actttctctt aaaattagca 2460
ggaagatggc cagtaaaagt aatacataca gacaatggca gcaatttcac cagtactaca 2520
gttaaggccg cctgttggtg ggcagggatc aagcaggaat ttggcattcc ctacaatccc 2580
caaagtcaag gagtagtaga atctatgaat aaagaattaa agaaaattat aggacaggta 2640
agagatcagg ctgaacatct taagacagca gtacaaatgg cagtattcat ccacaatttt 2700
aaaagaaaag gggggattgg ggggtacagt gcaggggaaa gaatagtaga cataatagca 2760
acagacatac aaactaaaga actacaaaaa caaattacaa aaattcaaaa ttttcgggtt 2820
tattacaggg acagcagaga tccactttgg aaaggaccag caaagcttct ctggaaaggt 2880
gaaggggcag tagtaataca agataatagt gacataaaag tagtgccaag aagaaaagca 2940
aagatcatta gggattatgg aaaacagatg gcaggtgatg attgtgtggc aagtagacag 3000
gatgaggatt ag 3012

<210> 33

<211> 3012

<212> DNA

<213> Artificial Sequence

<220>

<223> Least squares center of tree reconstruction of clade B pol gene s
equence

<400> 33
tttttaggg aagatctggc cttcccacaa gggaaggcca gggaattttc ttcagagcag 60

accagagcca acagccccac cagaagagag cttcaggttt ggggaagaga caacaactcc 120

ctctcagaag caggagccga tagacaagga actgtatcct ttagcttccc tcagatcact 180

ctttggcaac gacccctcgt cacaataaag ataggggggc aactaaagga agctctatta 240

gatacaggag cagatgatac agtattagaa gaaatgaatt tgccaggaag atggaaacca 300

aaaatgatag ggggaattgg aggttttatc aaagtaagac agtatgatca gatacccata 360

gaaatctgtg gacataaagc tataggtaca gtattagtag gacctacacc tgtcaacata 420

attggaagaa atctgttgac tcagattggt tgcactttaa attttcccat tagtcctatt 480

gaaactgtac cagtaaaatt aaagccagga atggatggcc caaaagttaa acaatggcca 540

ttgacagaag aaaaaataaa agcattagta gaaatttgta cagaaatgga aaaggaaggg 600

aaaatttcaa aaattgggcc tgaaaatcca tacaatactc cagtatttgc cataaagaaa 660

aaagacagta ctaaatggag aaaattagta gatttcagag aacttaataa gagaactcaa 720

gacttctggg aagttcaatt aggaatacca catcccgcag ggttaaaaaa gaaaaaatca 780

gtaacagtac tggatgtggg tgatgcatat ttttcagttc ccttagatga agacttcagg 840

aagtatactg catttaccat acctagtata aacaatgaga caccagggat tagatatcag 900

tacaatgtgc ttccacaggg atggaaagga tcaccagcaa tattccaaag tagcatgaca 960

aaaatcttag agccttttag aaaacaaaat ccagacatag ttatctatca atacatggat 1020

gatttgtatg taggatctga cttagaaata gggcagcata gaacaaaaat agaggaactg 1080

agacaacatc tgttgaggtg gggatttacc acaccagaca aaaaacatca gaaagaacct 1140

ccattccttt ggatgggtta tgaactccat cctgataaat ggacagtaca gcctatagtg 1200

ctgccagaaa aagacagctg gactgtcaat gacatacaga agttagtggg aaaattgaat 1260

tgggcaagtc agatttatgc agggattaaa gtaaagcaat tatgtaaact ccttagggga 1320

accaaagcac taacagaagt aataccacta acagaagaag cagagctaga actggcagaa 1380

aacagggaga ttctaaaaga accagtacat ggagtgtatt atgacccatc aaaagactta 1440

atagcagaaa tacagaagca ggggcaaggc caatggacat atcaaattta tcaagagcca 1500

tttaaaaatc tgaaaacagg aaagtatgca agaatgaggg gtgcccacac taatgatgta 1560

aaacaattaa cagaggcagt gcaaaaaata gccacagaaa gcatagtaat atggggaaag 1620

actcctaaat ttaaactacc catacaaaaa gaaacatggg aagcatggtg gacagagtat 1680

tggcaagcca cctggattcc tgagtgggag tttgtcaata cccctccctt agtgaaatta 1740

tggtaccagt tagagaaaga acccatagta ggagcagaaa ctttctatgt agatggggca 1800

gctaataggg agactaaatt aggaaaagca ggatatgtta ctgacagagg aagacaaaaa 1860

gttgtctccc taactgacac aacaaatcag aagactgagt tacaagcaat tcatctagct 1920

ttgcaggatt cgggattaga agtaaacata gtaacagact cacaatatgc attaggaatc 1980

attcaagcac aaccagataa gagtgaatca gagttagtca gtcaaataat agagcagtta 2040

ataaaaaagg aaaaggtcta cctggcatgg gtaccagcac acaaaggaat tggaggaaat 2100

gaacaagtag ataaattagt cagtgctgga atcaggaaag tactattttt ggatggaata 2160

gataaggccc aagaagaaca tgagaaatat cacagtaatt ggagagcaat ggctagtgat 2220

tttaacctgc cacctgtagt agcaaaagaa atagtagcca gctgtgataa atgtcagcta 2280

aaaggagaag ccatgcatgg acaagtagac tgtagtccag gaatatggca actagattgt 2340

acacatttag aaggaaaagt tatcctggta gcagttcatg tagccagtgg atatatagaa 2400

gcagaagtta ttccagcaga gacagggcag gaaacagcat actttctctt aaaattagca 2460

ggaagatggc cagtaaaaac aatacataca gacaatggca gcaatttcac cagtactacg 2520

gttaaggccg cctgttggtg ggcagggatc aagcaggaat ttggcattcc ctacaatccc 2580

caaagtcaag gagtagtaga atctatgaat aaagaattaa agaaaattat aggacaggta 2640

agagatcagg ctgaacatct taagacagca gtacaaatgg cagtattcat ccacaatttt 2700

aaaagaaaag gggggattgg ggggtacagt gcaggggaaa gaatagtaga cataatagca 2760

acagacatac aaactaaaga attacaaaaa caaattacaa aaattcaaaa ttttcgggtt 2820

tattacaggg acagcagaga tccactttgg aaaggaccag caaagcttct ctggaaaggt 2880

gaaggggcag tagtaataca agataatagt gacataaaag tagtgccaag aagaaaagca 2940

aagatcatta gggattatgg aaaacagatg gcaggtgatg attgtgtggc aagtagacag 3000

gatgaggatt ag 3012

<210> 34

<211> 3012

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade B pol gen
e sequence

<400> 34
tttttaggg aagatctggc cttcccacaa gggaaggcca gggaattttc ttcagagcag 60

accagagcca acagccccac cagaagagag cttcaggttt ggggaagaga caacaactcc 120

ctctcagaag caggagccga tagacaagga actgtatcct ttagcttccc tcagatcact 180

ctttggcaac gacccctcgt cacaataaag ataggggggc aactaaagga agctctatta 240

gatacaggag cagatgatac agtattagaa gaaatgaatt tgccaggaag atggaaacca 300

aaaatgatag ggggaattgg aggttttatc aaagtaagac agtatgatca gatactcata 360

gaaatctgtg gacataaagc tataggtaca gtattagtag gacctacacc tgtcaacata 420

attggaagaa atctgttgac tcagattggt tgcactttaa attttcccat tagtcctatt 480
gaaactgtac cagtaaaatt aaagccagga atggatggcc caaaagttaa acaatggcca 540
ttgacagaag aaaaaataaa agcattagta gaaatttgta cagaaatgga aaaggaaggg 600
aaaatttcaa aaattgggcc tgaaaatcca tacaatactc cagtatttgc cataaagaaa 660
aaagacagta ctaaatggag aaaattagta gatttcagag aacttaataa gagaactcaa 720
gacttctggg aagttcaatt aggaatacca catcccgcag ggttaaaaaa gaaaaaatca 780
gtaacagtac tggatgtggg tgatgcatat ttttcagttc ccttagatga agacttcagg 840
aagtatactg catttaccat acctagtata aacaatgaga caccagggat tagatatcag 900
tacaatgtgc ttccacaggg atggaaagga tcaccagcaa tattccaaag tagcatgaca 960
aaaatcttag agccttttag aaaacaaaat ccagacatag ttatctatca atacatggat 1020
gatttgtatg taggatctga cttagaaata gggcagcata gaacaaaaat agaggaactg 1080
agacaacatc tgttgaggtg gggatttacc acaccagaca aaaaacatca gaaagaacct 1140
ccattccttt ggatgggtta tgaactccat cctgataaat ggacagtaca gcctatagtg 1200
ctgccagaaa aagacagctg gactgtcaat gacatacaga agttagtggg aaaattgaat 1260
tgggcaagtc agatttaccc agggattaaa gtaaagcaat tatgtaaact ccttagggga 1320
accaaagcac taacagaagt aataccacta acagaagaag cagagctaga actggcagaa 1380
aacagggaaa ttctaaaaga accagtacat ggagtgtatt atgacccatc aaaagactta 1440
atagcagaaa tacagaagca ggggcaaggc caatggacat atcaaattta tcaagagcca 1500
tttaaaaatc tgaaaacagg aaaatatgca agaatgaggg gtgcccacac taatgatgta 1560
aaacaattaa cagaggcagt gcaaaaaata gccacagaaa gcatagtaat atggggaaag 1620
actcctaaat ttaaactacc catacaaaaa gaaacatggg aaacatggtg gacagagtat 1680
tggcaagcca cctggattcc tgagtgggag tttgtcaata cccctccctt agtgaaatta 1740
tggtaccagt tagagaaaga acccatagta ggagcagaaa ctttctatgt agatggggca 1800
gctaacaggg agactaaatt aggaaaagca ggatatgtta ctaacagagg aagacaaaaa 1860
gttgtctccc taactgacac aacaaatcag aagactgagt tacaagcaat tcatctagct 1920
ttgcaggatt cgggattaga agtaaacata gtaacagact cacaatatgc attaggaatc 1980
attcaagcac aaccagataa aagtgaatca gagttagtca gtcaaataat agagcagtta 2040
ataaaaaagg aaaaggtcta cctggcatgg gtaccagcac acaaaggaat tggaggaaat 2100
gaacaagtag ataaattagt cagtgctgga atcaggaaag tactattttt agatggaata 2160
gataaggccc aagaagaaca tgagaaatat cacagtaatt ggagagcaat ggctagtgat 2220
tttaacctgc cacctgtagt agcaaaagaa atagtagcca gctgtgataa atgtcagcta 2280
aaaggagaag ccatgcatgg acaagtagac tgtagtccag gaatatggca actagattgt 2340
acacatttag aaggaaaagt tatcctggta gcagttcatg tagccagtgg atatatagaa 2400
gcagaagtta ttccagcaga gacagggcag gaaacagcat actttctctt aaaattagca 2460

ggaagatggc cagtaaaaac aatacataca gacaatggca gcaatttcac cagtactacg 2520

gttaaggccg cctgttggtg ggcggggatc aagcaggaat ttggcattcc ctacaatccc 2580

caaagtcaag gagtagtaga atctatgaat aaagaattaa agaaaattat aggacaggta 2640

agagatcagg ctgaacatct taagacagca gtacaaatgg cagtattcat ccacaatttt 2700

aaaagaaaag gggggattgg ggggtacagt gcaggggaaa gaatagtaga cataatagca 2760

acagacatac aaactaaaga attacaaaaa caaattacaa aaattcaaaa ttttcgggtt 2820

tattacaggg acagcagaga tccactttgg aaaggaccag caaagcttct ctggaaaggt 2880

gaaggggcag tagtaataca agataatagt gacataaaag tagtgccaag aagaaaagca 2940

aagatcatta gggattatgg aaaacagatg gcaggtgatg attgtgtggc aagtagacag 3000

gatgaggatt ag 3012

<210> 35

<211> 360

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B rev gene se
quence

<400> 35
atggcaggaa gaagcggaga cagcgacgaa gagctcctca agacagtcag actcatcaag 60

tttctctatc aaagcaaccc gcctcccagc cccgagggga cccgacaggc ccgaaggaat 120

agaagaagaa ggtggagaga gagacagaga cagatccgtt cgattagtga acggattctt 180

agcacttatc tgggacgatc tgcggagcct gtgcctcttc agctaccacc gcttgagaga 240

cttactcttg attgtagcga ggattgtgga acttctggga cgcagggggt gggaagtcct 300

caaatattgg tggaatctcc tgcagtattg gagtcaggaa ctaaagaata gtgctgttag 360

<210> 36

<211> 360

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade B rev gene s
equence

<400> 36
atggcaggaa gaagcggaga cagcgacgaa gagctcctca agacagtcag actcatcaag 60

tttctctatc aaagcaaccc gcctcccagc cccgagggga cccgacaggc ccgaaggaat 120

cgaagaagaa ggtggagaga gagacagaga cagatccggt cgattagtga atggattctt 180

agcacttatc tgggtcgacc tgcggagcct gtgcctcttc agctaccacc gcttgagaga 240

cttactcttg attgtaacga ggattgtgga acttctggga cgcagggggt gggaagtcct 300

caaatattgg tggaatctcc tacagtattg gagtcaggaa ctaaagaata gtgctgttag 360

<210> 37

<211> 360

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade B rev gen
e sequence

<400> 37
atggcaggaa gaagcggaga cagcgacgaa gagctcctca agacagtcag actcatcaag 60

tttctctatc aaagcaaccc gcctcccagc cccgagggga cccgacaggc ccgaaggaat 120

cgaagaagaa ggtggagaga gagacagaga cagatccggt cgattagtga atggattctt 180

agcacttatc tgggacgacc tgcggagcct gtgcctcttc agctaccacc gcttgagaga 240

cttactcttg attgtagcga ggattgtgga acttctggga cgcagggggt gggaagtcct 300

caaatattgg tggaatctcc tgcagtattg gagtcaggaa ctaaagaata gtgctgttag 360

<210> 38

<211> 321

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B tat gene se
quence

<400> 38
atggagccag tagatcctag actagagccc tggaagcatc caggaagtca gcctaagact 60

gcttgtacca attgctattg taaaaagtgt tgctatcatt gccaagtttg cttcataaca 120

aaaggcttag gcatctccta tggcaggaag aagcggagac agcgacgaag acctcctcaa 180

ggcagtcaga ctcatcaagt ttctctatca aagcaacccg cctcccagcc ccgaggggac 240

ccgacaggcc cgaaggaatc gaagaagaag gtggagagag agacagagac agatccggtc 300

gattagtgaa tggattctta g 321

<210> 39

<211> 321

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade B tat gene s
equence

<400> 39
atggagccag tagatcctag actagagccc tggaagcatc caggaagtca gcctaagact 60

gcttgtacca attgctattg taaaaagtgt tgctttcatt gccaagtttg tttcataaca 120

aaaggcttag gcatctccta tggcaggaag aagcggagac agcgacgaag agctcctcaa 180

gacagtcaga ctcatcaagt ttctctatca aagcaacccg cctcccagcc ccgaggggac 240

ccgacaggcc cgaaggaatc gaagaagaag gtggagagag agacagagac agatccggtc 300

gattagtgaa tggattctta g 321

<210> 40

<211> 321

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade B tat gen
e sequence

<400> 40
atggagccag tagatcctag actagagccc tggaagcatc caggaagtca gcctaagact 60

gcttgtacca attgctattg taaaaagtgt tgctttcatt gccaagtttg tttcataaca 120

aaaggcttag gcatctccta tggcaggaag aagcggagac agcgacgaag agctcctcaa 180

gacagtcaga ctcatcaagt ttctctatca aagcaacccg cctcccagcc ccgaggggac 240

ccgacaggcc cgaaggaatc gaagaagaag gtggagagag agacagagac agatccggtc 300

gattagtgga tggattctta t 321

<210> 41

<211> 579

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B vif gene se
quence

<400> 41
atggaaaaca gatggcaggt gatgattgtg tggcaagtag acaggatgag gattagaaca 60

tggaaaagtt tagtaaaaca ccatatgtat atttcaaaga aagctaaggg atggttttat 120

agacatcact atgaaagcac tcatccaaga ataagttcag aagtacacat cccactagga 180

gatgctagat tggtaataaa aacatattgg ggtctgcata caggagaaag agaatggcat 240

ttgggtcagg gagtctccat agaatggagg aaaaggagat atagcacaca agtagaccct 300

ggcctagcag accaactaat tcatctgtat tattttgatt gtttttcaga atctgctata 360

agaaatgcca tattaggaca tatagttagt cctaggtgtg aatatcaagc aggacataac 420

aaggtaggat ctctacagta cttggcacta acagcattaa taacaccaaa aaagataaag 480

ccacctttgc ctagtgttag gaaactgaca gaggatagat ggaacaagcc ccagaagacc 540

aagggccaca gagggagcca tacaatgaat ggacactag 579

<210> 42

<211> 579

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares and minimum of means center of tree reconstructions
of clade B vif gene sequence

<400> 42
atggaaaaca gatggcaggt gatgattgtg tggcaagtag acaggatgag gattagaaca 60

tggaaaagtt tagtaaaaca ccatatgtat atttcaagga aagctaaggg atggttttat 120

agacatcact atgaaagcac tcatccaaga ataagttcag aagtacacat cccactaggg 180

gatgctagat tggtaataac aacatattgg ggtctgcata caggagaaag agactggcat 240

ttgggtcagg gagtctccat agaatggagg aaaaagagat atagcacaca agtagaccct 300

gacctagcag accaactaat tcatctgtat tactttgatt gtttttcaga atctgctata 360

agaaatgcca tattaggaca tatagttagt cctaggtgtg aatatcaagc aggacataac 420

aaggtaggat ctctacagta cttggcacta gcagcattaa taacaccaaa aaagataaag 480

ccacctttgc ctagtgttac gaaactgaca gaggatagat ggaacaagcc ccagaagacc 540

aagggccaca gagggagcca tacaatgaat ggacactag 579

<210> 43

<211> 291

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade b vpr gene se
quence

<400> 43
atggaacaag ccccagaaga ccaagggcca cagagggagc catacaatga atggacacta 60

gagcttttag aggagcttaa gagtgaagct gttagacatt ttcctaggct atggctccat 120

agcttaggac aacatatcta tgaaacttat ggggatacct gggcaggagt ggaagctata 180

ataagaattc tgcaacaact gctgtttatt catttcagaa ttgggtgtca acatagcaga 240

ataggcatta ctcgacagag aagagcaaga aatggagcca gtagatccta g 291

<210> 44

<211> 291

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares and minimum of means center of tree reconstruction
of clade B vpr gene sequence

<400> 44
atggaacaag ccccagaaga ccaagggcca cagagggagc catacaatga atggacacta 60

gagcttttag aggagcttaa gagtgaagct gttagacatt ttcctaggat atggctccat 120

agcttaggac aacatatcta tgaaacttat ggggatactt gggcaggagt ggaagccata 180

ataagaattc tgcaacaact gctgtttatt catttcagaa ttgggtgtcg acatagcaga 240

ataggcatta ctcgacagag gagagcaaga aatggagcca gtagatccta g 291

<210> 45

<211> 246

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B vpu gene se
quence

<400> 45

atgcaacctt tagaaatatt agcaatagta gcattagtag tagcagcaat actagcaata 60

gttgtgtgga ccatagtatt catagaatat aggaaaatat taaggcaaag aaaaatagac 120

aggttaattg atagaataag agaaagagca gaagacagtg gcaatgagag tgaaggggat 180

caggaagaat tatcagcact tgtggaaatg gggcaccatg ctccttggga tgttgatgat 240

ctgtag 246

<210> 46

<211> 246

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares and minimum of means center of tree reconstructions of clade B vpu gene sequence

<400> 46
atgcaacctt tacaaatatt agcaatagta gcattagtag tagcagcaat aatagcaata 60

gttgtgtgga ccatagtatt catagaatat aggaaaatat taagacaaag aaaaatagac 120

aggttaattg atagaataag agaaagagca gaagacagtg gcaatgagag tgaaggggat 180

caggaagaat tatcagcact tgtggagatg gggcaccatg ctccttggga tgttgatgat 240

ctgtag 246

<210> 47

<211> 500

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B gag protein sequence

<400> 47

Met Gly Ala Arg Ala Ser Val Leu Ser Gly Gly Glu Leu Asp Lys Trp
1 5 10 15

Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys Lys Tyr Lys Leu Lys
20 25 30

His Ile Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Val Asn Pro
35 40 45

Gly Leu Leu Glu Thr Ser Glu Gly Cys Arg Gln Ile Leu Gly Gln Leu

```
    50                  55                  60

Gln Pro Ser Leu Gln Thr Gly Ser Glu Glu Leu Arg Ser Leu Tyr Asn
65                  70                  75                  80

Thr Val Ala Val Leu Tyr Cys Val His Gln Lys Ile Glu Val Lys Asp
                85                  90                  95

Thr Lys Glu Ala Leu Asp Lys Ile Glu Glu Glu Gln Asn Lys Ser Lys
            100                 105                 110

Lys Lys Ala Gln Gln Ala Ala Ala Asp Thr Gly Asn Ser Ser Gln Val
        115                 120                 125

Ser Gln Asn Tyr Pro Ile Val Gln Asn Leu Gln Gly Gln Met Val His
    130                 135                 140

Gln Ala Leu Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Ile Glu
145                 150                 155                 160

Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Ser Ala Leu Ser
                165                 170                 175

Glu Gly Ala Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly
            180                 185                 190

Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile Asn Glu Glu
        195                 200                 205

Ala Ala Glu Trp Asp Arg Leu His Pro Val His Ala Gly Pro Ile Ala
    210                 215                 220

Pro Gly Gln Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr
225                 230                 235                 240

Ser Thr Leu Gln Glu Gln Ile Ala Trp Met Thr Asn Asn Pro Pro Ile
                245                 250                 255

Pro Val Gly Glu Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys
            260                 265                 270

Ile Val Arg Met Tyr Ser Pro Val Ser Ile Leu Asp Ile Arg Gln Gly
        275                 280                 285

Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Tyr Lys Thr Leu
    290                 295                 300

Arg Ala Glu Gln Ala Ser Gln Glu Val Lys Asn Trp Met Thr Glu Thr
305                 310                 315                 320

Leu Leu Val Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Lys Ala
```

325 330 335

Leu Gly Pro Gly Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly
340 345 350

Val Gly Gly Pro Gly His Lys Ala Arg Val Leu Ala Glu Ala Met Ser
355 360 365

Gln Val Thr Asn Ser Ala Thr Ile Met Met Gln Arg Gly Asn Phe Arg
370 375 380

Asn Pro Arg Lys Thr Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His
385 390 395 400

Ile Ala Arg Asn Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys
405 410 415

Gly Lys Glu Gly His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn
420 425 430

Phe Leu Gly Lys Ile Trp Pro Ser His Lys Gly Arg Pro Gly Asn Phe
435 440 445

Leu Gln Ser Arg Pro Glu Pro Thr Ala Pro Pro Glu Glu Ser Phe Arg
450 455 460

Phe Gly Glu Glu Thr Thr Thr Pro Ser Gln Lys Gln Glu Gln Lys Asp
465 470 475 480

Lys Glu Leu Tyr Pro Leu Ala Ser Leu Lys Ser Leu Phe Gly Asn Asp
485 490 495

Pro Ser Ser Gln
500

<210> 48

<211> 500

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade B gag protei
n sequence

<400> 48

Met Gly Ala Arg Ala Ser Val Leu Ser Gly Gly Glu Leu Asp Arg Trp
1 5 10 15

Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys Lys Tyr Arg Leu Lys
20 25 30

His Ile Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Val Asn Pro
35 40 45

Gly Leu Leu Glu Thr Ser Glu Gly Cys Arg Gln Ile Leu Gly Gln Leu
50 55 60

Gln Pro Ser Leu Gln Thr Gly Ser Glu Glu Leu Arg Ser Leu Tyr Asn
65 70 75 80

Thr Val Ala Thr Leu Tyr Cys Val His Gln Arg Ile Glu Val Lys Asp
85 90 95

Thr Lys Glu Ala Leu Glu Lys Ile Glu Glu Glu Gln Asn Lys Ser Lys
100 105 110

Lys Lys Ala Gln Gln Ala Ala Ala Asp Thr Gly Asn Ser Ser Gln Val
115 120 125

Ser Gln Asn Tyr Pro Ile Val Gln Asn Leu Gln Gly Gln Met Val His
130 135 140

Gln Ala Ile Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Val Glu
145 150 155 160

Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Ser Ala Leu Ser
165 170 175

Glu Gly Ala Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly
180 185 190

Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile Asn Glu Glu
195 200 205

Ala Ala Glu Trp Asp Arg Leu His Pro Val His Ala Gly Pro Ile Ala
210 215 220

Pro Gly Gln Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr
225 230 235 240

Ser Thr Leu Gln Glu Gln Ile Gly Trp Met Thr Asn Asn Pro Pro Ile
245 250 255

Pro Val Gly Glu Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys
260 265 270

Ile Val Arg Met Tyr Ser Pro Thr Ser Ile Leu Asp Ile Arg Gln Gly
275 280 285

Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Tyr Lys Thr Leu
290 295 300

Arg Ala Glu Gln Ala Ser Gln Glu Val Lys Asn Trp Met Thr Glu Thr
305 310 315 320

Leu Leu Val Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Lys Ala
325 330 335

Leu Gly Pro Ala Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly
340 345 350

Val Gly Gly Pro Gly His Lys Ala Arg Val Leu Ala Glu Ala Met Ser
355 360 365

Gln Val Thr Asn Ser Ala Thr Ile Met Met Gln Arg Gly Asn Phe Arg
370 375 380

Asn Gln Arg Lys Thr Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His
385 390 395 400

Ile Ala Lys Asn Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys
405 410 415

Gly Lys Glu Gly His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn
420 425 430

Phe Leu Gly Lys Ile Trp Pro Ser His Lys Gly Arg Pro Gly Asn Phe
435 440 445

Leu Gln Ser Arg Pro Glu Pro Thr Ala Pro Pro Glu Glu Ser Phe Arg
450 455 460

Phe Gly Glu Glu Thr Thr Thr Pro Ser Gln Lys Gln Glu Pro Ile Asp
465 470 475 480

Lys Glu Leu Tyr Pro Leu Ala Ser Leu Arg Ser Leu Phe Gly Asn Asp
485 490 495

Pro Ser Ser Gln
500

<210> 49

<211> 500

<212> PRT

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade B gag pro
tein sequence

<400> 49

Met Gly Ala Gly Ala Ser Val Leu Ser Gly Gly Lys Leu Asp Arg Trp
1   5   10   15

Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys Lys Tyr Lys Leu Lys
            20   25   30

His Ile Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Val Asn Pro
        35   40   45

Gly Leu Leu Glu Thr Ser Glu Gly Cys Arg Arg Ile Leu Glu Gln Leu
    50   55   60

His Pro Ser Leu Gln Thr Gly Ser Glu Glu Leu Lys Ser Leu Tyr Asn
65   70   75   80

Thr Val Ala Thr Leu Tyr Cys Val His Gln Asn Ile Glu Val Arg Asp
                85   90   95

Thr Lys Asp Ala Leu Glu Lys Ile Glu Glu Glu Gln Asn Lys Ile Lys
            100   105   110

Lys Arg Ala Gln Gln Ala Ala Ala Asp Thr Gly Asn Ser Asn Pro Val
        115   120   125

Ser Gln Asn Tyr Pro Ile Val Gln Asn Met Gln Gly Gln Met Val His
    130   135   140

Gln Ala Ile Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Val Glu
145   150   155   160

Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Ser Ala Leu Ser
                165   170   175

Glu Gly Ala Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly
            180   185   190

Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile Asn Glu Glu
        195   200   205

Ala Ala Glu Trp Asp Arg Leu His Pro Val His Ala Gly Pro Ile Ala
    210   215   220

Pro Gly Gln Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr
225   230   235   240

Ser Thr Leu Gln Glu Gln Ile Gly Trp Met Thr His Asn Pro Pro Ile
                245   250   255

Pro Val Gly Glu Ile Tyr Lys Arg Trp Ile Ile Met Gly Leu Asn Lys
260 265 270

Ile Val Arg Met Tyr Ser Pro Thr Ser Ile Leu Asp Ile Arg Gln Gly
275 280 285

Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Tyr Lys Thr Leu
290 295 300

Arg Ala Glu Gln Ala Ser Gln Glu Val Lys Asn Trp Met Thr Glu Thr
305 310 315 320

Leu Leu Val Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Lys Ala
325 330 335

Leu Gly Pro Ala Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly
340 345 350

Val Gly Gly Pro Ser His Lys Ala Arg Val Leu Ala Glu Ala Met Ser
355 360 365

Gln Ala Thr Asn Ser Ala Thr Ile Met Met Gln Arg Gly Asn Phe Lys
370 375 380

Gly Gln Arg Lys Thr Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His
385 390 395 400

Ile Ala Arg Asn Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys
405 410 415

Gly Lys Glu Gly His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn
420 425 430

Phe Leu Gly Lys Ile Trp Pro Ser His Lys Gly Arg Pro Gly Asn Phe
435 440 445

Leu Gln Ser Arg Pro Glu Pro Thr Ala Pro Pro Glu Glu Ser Phe Arg
450 455 460

Phe Gly Glu Glu Thr Thr Thr Pro Pro Gln Lys Gln Glu Pro Arg Asp
465 470 475 480

Lys Glu Gln Tyr Pro Leu Thr Ser Leu Arg Ser Leu Phe Gly Asn Asp
485 490 495

Pro Ser Ser Gln
500

<210> 50

<211> 862

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B gp 160 prot
ein sequence

<400> 50

```
Met Arg Val Lys Gly Ile Arg Lys Asn Cys Gln His Leu Trp Lys Trp
1               5                   10                  15

Gly Thr Met Leu Leu Gly Met Leu Met Ile Cys Ser Ala Ala Glu Asn
            20                  25                  30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr
        35                  40                  45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Lys Thr Glu Val
    50                  55                  60

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65                  70                  75                  80

Gln Glu Val Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95

Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
            100                 105                 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125

Asn Cys Thr Asp Ala Asn Lys Asn Ala Thr Asn Thr Asn Ser Ser Ser
    130                 135                 140

Gly Gly Thr Met Glu Lys Gly Glu Met Lys Asn Cys Ser Phe Asn Ile
145                 150                 155                 160

Thr Thr Ser Ile Arg Asp Lys Met Gln Lys Glu Tyr Ala Leu Phe Tyr
                165                 170                 175

Lys Leu Asp Val Val Pro Ile Asp Asn Asp Asn Asn Ser Asn Asn Asn
            180                 185                 190

Thr Asn Tyr Arg Leu Ile Asn Cys Asn Thr Ser Val Ile Thr Gln Ala
        195                 200                 205

Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Thr Pro
    210                 215                 220
```

Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn Gly Thr
225 230 235 240

Gly Pro Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg
245 250 255

Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu
260 265 270

Glu Val Val Ile Arg Ser Glu Asn Phe Thr Asp Asn Ala Lys Thr Ile
275 280 285

Ile Val Gln Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn
290 295 300

Asn Asn Thr Arg Lys Ser Ile Pro Ile Gly Pro Gly Arg Ala Leu Tyr
305 310 315 320

Thr Thr Gly Glu Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Ile
325 330 335

Ser Arg Ala Lys Trp Asn Asn Thr Leu Lys Gln Val Val Thr Lys Leu
340 345 350

Arg Glu Gln Phe Gly Asn Asn Lys Thr Ile Val Phe Asn Pro Ser Ser
355 360 365

Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu
370 375 380

Phe Phe Tyr Cys Asn Thr Thr Gln Leu Phe Asn Ser Thr Trp Asn Ser
385 390 395 400

Thr Glu Gly Ser Asn Lys Thr Thr Gly Ser Asn Asn Thr Gly Gly Glu
405 410 415

Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln
420 425 430

Glu Val Gly Lys Ala Met Tyr Ala Pro Pro Ile Arg Gly Gln Ile Lys
435 440 445

Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Glu
450 455 460

Asn Ser Thr Asn Glu Thr Glu Ile Phe Arg Pro Gly Gly Gly Asp Met
465 470 475 480

Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile
485 490 495

Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln
            500   505   510

Arg Glu Lys Arg Ala Val Gly Ile Ile Gly Ala Met Phe Leu Gly Phe
        515   520   525

Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Met Thr Leu Thr
    530   535   540

Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn
545   550   555   560

Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val
                565   570   575

Trp Gly Ile Lys Gln Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr
            580   585   590

Leu Arg Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu
        595   600   605

Ile Cys Thr Thr Thr Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser
    610   615   620

Leu Asp Lys Ile Trp Asn Asn Met Thr Trp Met Glu Trp Glu Arg Glu
625   630   635   640

Ile Asp Asn Tyr Thr Gly Leu Ile Tyr Asn Leu Ile Glu Glu Ser Gln
                645   650   655

Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp
            660   665   670

Ala Ser Leu Trp Asn Trp Phe Asp Ile Thr Gln Trp Leu Trp Tyr Ile
        675   680   685

Lys Ile Phe Ile Met Ile Val Gly Gly Leu Val Gly Leu Arg Ile Val
    690   695   700

Phe Ala Val Leu Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro
705   710   715   720

Leu Ser Phe Gln Thr Arg Leu Pro Ala Pro Arg Gly Pro Asp Arg Pro
                725   730   735

Glu Gly Ile Glu Glu Glu Gly Gly Glu Arg Asp Arg Asp Arg Ser Gly
            740   745   750

Arg Leu Val Asn Gly Phe Leu Ala Leu Ile Trp Asp Asp Leu Arg Ser
        755   760   765

Leu Cys Leu Phe Ser Tyr His Arg Leu Arg Asp Leu Leu Leu Ile Val
    770   775   780

Ala Arg Ile Val Glu Leu Leu Gly Arg Arg Gly Trp Glu Ala Leu Lys
785   790   795   800

Tyr Trp Trp Asn Leu Leu Gln Tyr Trp Ser Gln Glu Leu Lys Asn Ser
                805   810   815

Ala Val Ser Leu Leu Asn Ala Thr Ala Ile Ala Val Ala Glu Gly Thr
            820   825   830

Asp Arg Val Ile Glu Val Val Gln Arg Ala Cys Arg Ala Ile Leu His
        835   840   845

Ile Pro Arg Arg Ile Arg Gln Gly Leu Glu Arg Ala Leu Leu
    850   855   860

<210> 51

<211> 862

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares and minimum of means center of tree reconstruction of clade B gp 160 protein sequence

<400> 51

Met Arg Val Lys Gly Ile Arg Lys Asn Tyr Gln His Leu Trp Arg Trp
1   5   10   15

Gly Thr Met Leu Leu Gly Met Leu Met Ile Cys Ser Ala Ala Glu Lys
            20   25   30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr
        35   40   45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val
    50   55   60

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65   70   75   80

Gln Glu Val Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85   90   95

Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp

100 105 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
115 120 125

Asn Cys Thr Asp Leu Asn Lys Asn Ala Thr Asn Thr Asn Ser Ser Ser
130 135 140

Gly Glu Met Met Glu Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn Ile
145 150 155 160

Thr Thr Ser Ile Arg Asp Lys Val Gln Lys Glu Tyr Ala Leu Phe Tyr
165 170 175

Lys Leu Asp Val Val Pro Ile Asp Asn Asp Asn Asn Thr Asn Asn Thr
180 185 190

Thr Ser Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln Ala
195 200 205

Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro
210 215 220

Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn Gly Thr
225 230 235 240

Gly Pro Cys Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg
245 250 255

Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu
260 265 270

Glu Val Val Ile Arg Ser Asp Asn Phe Thr Asp Asn Ala Lys Thr Ile
275 280 285

Ile Val Gln Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn
290 295 300

Asn Asn Thr Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe Tyr
305 310 315 320

Thr Thr Gly Glu Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Ile
325 330 335

Ser Arg Ala Lys Trp Asn Asn Thr Leu Lys Gln Ile Val Lys Lys Leu
340 345 350

Arg Glu Gln Phe Gly Asn Asn Lys Thr Ile Val Phe Asn Gln Ser Ser
355 360 365

Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu

370 375 380

Phe Phe Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp Asn Gly
385 390 395 400

Thr Trp Thr Trp Asn Thr Thr Glu Gly Ser Asn Asp Thr Glu Gly Asp
405 410 415

Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln
420 425 430

Glu Val Gly Lys Ala Met Tyr Ala Pro Pro Ile Arg Gly Gln Ile Arg
435 440 445

Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Asn
450 455 460

Asn Asn Thr Asn Glu Thr Glu Ile Phe Arg Pro Gly Gly Gly Asp Met
465 470 475 480

Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile
485 490 495

Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln
500 505 510

Arg Glu Lys Arg Ala Val Gly Ile Ile Gly Ala Val Phe Leu Gly Phe
515 520 525

Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Met Thr Leu Thr
530 535 540

Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn
545 550 555 560

Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val
565 570 575

Trp Gly Ile Lys Gln Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr
580 585 590

Leu Arg Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu
595 600 605

Ile Cys Thr Thr Ala Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser
610 615 620

Leu Asp Glu Ile Trp Asn Asn Met Thr Trp Met Glu Trp Glu Arg Glu
625 630 635 640

Ile Asp Asn Tyr Thr Ser Leu Ile Tyr Thr Leu Ile Glu Glu Ser Gln

645 650 655

Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp
660 665 670

Ala Ser Leu Trp Asn Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile
675 680 685

Lys Ile Phe Ile Met Ile Val Gly Gly Leu Val Gly Leu Arg Ile Val
690 695 700

Phe Ala Val Leu Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro
705 710 715 720

Leu Ser Phe Gln Thr Arg Leu Pro Ala Pro Arg Gly Pro Asp Arg Pro
725 730 735

Glu Gly Ile Glu Glu Glu Gly Gly Glu Arg Asp Arg Asp Arg Ser Gly
740 745 750

Arg Leu Val Asn Gly Phe Leu Ala Leu Ile Trp Asp Asp Leu Arg Ser
755 760 765

Leu Cys Leu Phe Ser Tyr His Arg Leu Arg Asp Leu Leu Leu Ile Val
770 775 780

Thr Arg Ile Val Glu Leu Leu Gly Arg Arg Gly Trp Glu Ala Leu Lys
785 790 795 800

Tyr Trp Trp Asn Leu Leu Gln Tyr Trp Ser Gln Glu Leu Lys Asn Ser
805 810 815

Ala Val Ser Leu Leu Asn Ala Thr Ala Ile Ala Val Ala Glu Gly Thr
820 825 830

Asp Arg Val Ile Glu Val Val Gln Arg Ala Cys Arg Ala Ile Leu His
835 840 845

Ile Pro Thr Arg Ile Arg Gln Gly Leu Glu Arg Ala Leu Leu
850 855 860

<210> 52

<211> 206

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B nef protein

sequence

<400> 52

Met Gly Gly Lys Trp Ser Lys Arg Ser Val Val Gly Trp Pro Ala Val
1 5 10 15

Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Gly Val Gly Ala
20 25 30

Val Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr
35 40 45

Ala Ala Thr Asn Ala Ala Cys Ala Trp Leu Glu Ala Gln Glu Glu Glu
50 55 60

Glu Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr
65 70 75 80

Tyr Lys Ala Ala Val Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly
85 90 95

Leu Glu Gly Leu Val Tyr Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu
100 105 110

Trp Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr
115 120 125

Pro Gly Pro Gly Thr Arg Tyr Pro Leu Thr Phe Gly Trp Cys Phe Lys
130 135 140

Leu Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Thr Glu Gly Glu
145 150 155 160

Asn Asn Ser Leu Leu His Pro Met Ser Leu His Gly Met Asp Asp Pro
165 170 175

Glu Arg Glu Val Leu Val Trp Arg Phe Asp Ser Arg Leu Ala Phe His
180 185 190

His Met Ala Arg Glu Lys His Pro Glu Tyr Tyr Lys Asp Cys
195 200 205

<210> 53

<211> 206

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade B nef protei
n sequence

<400> 53

Met Gly Gly Lys Trp Ser Lys Arg Ser Val Val Gly Trp Pro Ala Val
1   5   10   15

Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Gly Val Gly Ala
            20   25   30

Val Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr
        35   40   45

Ala Ala Thr Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Glu Glu
    50   55   60

Glu Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr
65   70   75   80

Tyr Lys Ala Ala Leu Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly
                85   90   95

Leu Glu Gly Leu Ile Tyr Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu
            100   105   110

Trp Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr
        115   120   125

Pro Gly Pro Gly Ile Arg Tyr Pro Leu Thr Phe Gly Trp Cys Phe Lys
    130   135   140

Leu Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu
145   150   155   160

Asn Asn Ser Leu Leu His Pro Met Ser Leu His Gly Met Asp Asp Pro
                165   170   175

Glu Lys Glu Val Leu Val Trp Lys Phe Asp Ser Arg Leu Ala Phe His
            180   185   190

His Met Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys
        195   200   205

<210> 54

<211> 206

<212> PRT

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade B nef pro tein sequence

<400> 54

Met Gly Gly Lys Trp Ser Lys Arg Ser Val Val Gly Trp Pro Ala Val
1   5   10   15

Arg Glu Arg Met Arg Arg Ala Glu Pro Ala Ala Asp Gly Val Gly Ala
            20   25   30

Val Ser Arg Asp Leu Glu Lys His Gly Ala Ile Thr Ser Ser Asn Thr
        35   40   45

Ala Ala Thr Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Glu Glu
    50   55   60

Glu Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr
65   70   75   80

Tyr Lys Ala Ala Leu Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly
                85   90   95

Leu Glu Gly Leu Ile Tyr Ser Gln Lys Arg Gln Asp Ile Leu Asp Leu
            100   105   110

Trp Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr
        115   120   125

Pro Gly Pro Gly Ile Arg Tyr Pro Leu Thr Phe Gly Trp Cys Phe Lys
    130   135   140

Leu Val Pro Val Glu Pro Glu Lys Val Glu Glu Ala Asn Glu Gly Glu
145   150   155   160

Asn Asn Cys Leu Leu His Pro Met Ser Gln His Gly Met Asp Asp Pro
                165   170   175

Glu Lys Glu Val Leu Val Trp Lys Phe Asp Ser Arg Leu Ala Phe His
            180   185   190

His Met Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys
        195   200   205

<210> 55

<211> 1003

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B pol protein
 sequence

<400> 55

Phe Phe Arg Glu Asn Leu Ala Phe Pro Gln Gly Lys Ala Arg Glu Leu
1   5   10   15

Ser Ser Glu Gln Thr Arg Ala Asn Ser Pro Thr Arg Arg Glu Leu Gln
            20   25   30

Val Trp Gly Arg Asp Asn Asn Ser Leu Ser Glu Ala Gly Ala Asp Arg
        35   40   45

Gln Gly Thr Val Ser Phe Ser Phe Pro Gln Ile Thr Leu Trp Gln Arg
    50   55   60

Pro Leu Val Thr Ile Lys Ile Gly Gly Gln Leu Lys Glu Ala Leu Leu
65   70   75   80

Asp Thr Gly Ala Asp Asp Thr Val Leu Glu Glu Met Asn Leu Pro Gly
                85   90   95

Lys Trp Lys Pro Lys Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val
            100   105   110

Arg Gln Tyr Asp Gln Ile Pro Ile Glu Ile Cys Gly His Lys Ala Ile
        115   120   125

Gly Thr Val Leu Val Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn
    130   135   140

Leu Leu Thr Gln Ile Gly Cys Thr Leu Asn Phe Pro Ile Ser Pro Ile
145   150   155   160

Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val
                165   170   175

Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile
            180   185   190

Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu
        195   200   205

Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr
    210   215   220

Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln
225   230   235   240

Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys

245 250 255

Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser
260 265 270

Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro
275 280 285

Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu
290 295 300

Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr
305 310 315 320

Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Glu Ile Val Ile Tyr
325 330 335

Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln
340 345 350

His Arg Thr Lys Ile Glu Glu Leu Arg Glu His Leu Leu Arg Trp Gly
355 360 365

Phe Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp
370 375 380

Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val
385 390 395 400

Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val
405 410 415

Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Ala Gly Ile Lys Val Lys
420 425 430

Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Val
435 440 445

Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile
450 455 460

Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu
465 470 475 480

Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile
485 490 495

Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met
500 505 510

Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln

515 520 525

Lys Ile Ala Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe
530 535 540

Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Ala Trp Trp Thr Glu Tyr
545 550 555 560

Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro
565 570 575

Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala
580 585 590

Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly
595 600 605

Lys Ala Gly Tyr Val Thr Asp Arg Gly Arg Gln Lys Val Val Ser Leu
610 615 620

Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile His Leu Ala
625 630 635 640

Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr
645 650 655

Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Lys Ser Glu Ser Glu Leu
660 665 670

Val Ser Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu
675 680 685

Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp
690 695 700

Lys Leu Val Ser Thr Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile
705 710 715 720

Asp Lys Ala Gln Glu Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala
725 730 735

Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val
740 745 750

Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln
755 760 765

Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu
770 775 780

Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu

785 790 795 800

Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu
805 810 815

Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Val Ile His Thr Asp Asn
820 825 830

Gly Ser Asn Phe Thr Ser Thr Thr Val Lys Ala Ala Cys Trp Trp Ala
835 840 845

Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly
850 855 860

Val Val Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val
865 870 875 880

Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe
885 890 895

Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly
900 905 910

Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu
915 920 925

Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp
930 935 940

Ser Arg Asp Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly
945 950 955 960

Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro
965 970 975

Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly
980 985 990

Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp
995 1000

<210> 56

<211> 1003

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade B pol protei

n sequence

<400> 56

Phe Phe Arg Glu Asp Leu Ala Phe Pro Gln Gly Lys Ala Arg Glu Phe
1   5   10   15

Ser Ser Glu Gln Thr Arg Ala Asn Ser Pro Thr Arg Arg Glu Leu Gln
            20   25   30

Val Trp Gly Arg Asp Asn Asn Ser Leu Ser Glu Ala Gly Ala Asp Arg
        35   40   45

Gln Gly Thr Val Ser Phe Ser Phe Pro Gln Ile Thr Leu Trp Gln Arg
    50   55   60

Pro Leu Val Thr Ile Lys Ile Gly Gly Gln Leu Lys Glu Ala Leu Leu
65   70   75   80

Asp Thr Gly Ala Asp Asp Thr Val Leu Glu Glu Met Asn Leu Pro Gly
                85   90   95

Arg Trp Lys Pro Lys Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val
            100   105   110

Arg Gln Tyr Asp Gln Ile Pro Ile Glu Ile Cys Gly His Lys Ala Ile
        115   120   125

Gly Thr Val Leu Val Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn
    130   135   140

Leu Leu Thr Gln Ile Gly Cys Thr Leu Asn Phe Pro Ile Ser Pro Ile
145   150   155   160

Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val
                165   170   175

Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile
            180   185   190

Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu
        195   200   205

Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr
    210   215   220

Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln
225   230   235   240

Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys
                245   250   255

Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser
260 265 270

Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro
275 280 285

Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu
290 295 300

Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr
305 310 315 320

Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr
325 330 335

Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln
340 345 350

His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly
355 360 365

Phe Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp
370 375 380

Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val
385 390 395 400

Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val
405 410 415

Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Ala Gly Ile Lys Val Lys
420 425 430

Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile
435 440 445

Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile
450 455 460

Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu
465 470 475 480

Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile
485 490 495

Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met
500 505 510

Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln
515 520 525

Lys Ile Ala Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe
    530   535   540

Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Ala Trp Trp Thr Glu Tyr
545   550   555   560

Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro
                565   570   575

Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala
            580   585   590

Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly
        595   600   605

Lys Ala Gly Tyr Val Thr Asp Arg Gly Arg Gln Lys Val Val Ser Leu
    610   615   620

Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile His Leu Ala
625   630   635   640

Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr
                645   650   655

Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Lys Ser Glu Ser Glu Leu
            660   665   670

Val Ser Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu
        675   680   685

Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp
    690   695   700

Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile
705   710   715   720

Asp Lys Ala Gln Glu Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala
                725   730   735

Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val
            740   745   750

Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln
        755   760   765

Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu
    770   775   780

Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu
785   790   795   800

Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu
805 810 815

Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn
820 825 830

Gly Ser Asn Phe Thr Ser Thr Thr Val Lys Ala Ala Cys Trp Trp Ala
835 840 845

Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly
850 855 860

Val Val Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val
865 870 875 880

Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe
885 890 895

Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly
900 905 910

Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu
915 920 925

Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp
930 935 940

Ser Arg Asp Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly
945 950 955 960

Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro
965 970 975

Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly
980 985 990

Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp
995 1000

<210> 57

<211> 1003

<212> PRT

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade B pol pro
tein sequence

<400> 57

Phe Phe Arg Glu Asp Leu Ala Phe Pro Gln Gly Lys Ala Arg Glu Phe
1   5   10   15

Ser Ser Glu Gln Thr Arg Ala Asn Ser Pro Thr Arg Arg Glu Leu Gln
            20   25   30

Val Trp Gly Arg Asp Asn Asn Ser Leu Ser Glu Ala Gly Ala Asp Arg
        35   40   45

Gln Gly Thr Val Ser Phe Ser Phe Pro Gln Ile Thr Leu Trp Gln Arg
    50   55   60

Pro Leu Val Thr Ile Lys Ile Gly Gly Gln Leu Lys Glu Ala Leu Leu
65   70   75   80

Asp Thr Gly Ala Asp Asp Thr Val Leu Glu Glu Met Asn Leu Pro Gly
                85   90   95

Arg Trp Lys Pro Lys Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val
            100   105   110

Arg Gln Tyr Asp Gln Ile Leu Ile Glu Ile Cys Gly His Lys Ala Ile
        115   120   125

Gly Thr Val Leu Val Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn
    130   135   140

Leu Leu Thr Gln Ile Gly Cys Thr Leu Asn Phe Pro Ile Ser Pro Ile
145   150   155   160

Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val
                165   170   175

Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile
            180   185   190

Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu
        195   200   205

Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr
    210   215   220

Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln
225   230   235   240

Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys
                245   250   255

Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser
            260   265   270

```
Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro
        275                 280                 285

Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu
    290                 295                 300

Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr
305                 310                 315                 320

Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr
                325                 330                 335

Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln
            340                 345                 350

His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly
        355                 360                 365

Phe Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp
    370                 375                 380

Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val
385                 390                 395                 400

Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val
                405                 410                 415

Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Lys
            420                 425                 430

Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile
        435                 440                 445

Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile
    450                 455                 460

Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu
465                 470                 475                 480

Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile
                485                 490                 495

Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met
            500                 505                 510

Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln
        515                 520                 525

Lys Ile Ala Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe
    530                 535                 540
```

Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr
545 550 555 560

Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro
565 570 575

Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala
580 585 590

Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly
595 600 605

Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Ser Leu
610 615 620

Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile His Leu Ala
625 630 635 640

Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr
645 650 655

Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Lys Ser Glu Ser Glu Leu
660 665 670

Val Ser Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu
675 680 685

Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp
690 695 700

Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile
705 710 715 720

Asp Lys Ala Gln Glu Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala
725 730 735

Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val
740 745 750

Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln
755 760 765

Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu
770 775 780

Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu
785 790 795 800

Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu
805 810 815

Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn
            820   825   830

Gly Ser Asn Phe Thr Ser Thr Thr Val Lys Ala Ala Cys Trp Trp Ala
        835   840   845

Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly
    850   855   860

Val Val Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val
865   870   875   880

Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe
                885   890   895

Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly
            900   905   910

Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu
        915   920   925

Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp
    930   935   940

Ser Arg Asp Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly
945   950   955   960

Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro
                965   970   975

Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly
            980   985   990

Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp
        995   1000

<210> 58

<211> 116

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B rev protein
       sequence

<400> 58

Met Ala Gly Arg Ser Gly Asp Ser Asp Glu Glu Leu Leu Lys Thr Val
1   5   10   15

Arg Leu Ile Lys Phe Leu Tyr Gln Ser Asn Pro Pro Pro Ser Pro Glu
20 25 30

Gly Thr Arg Gln Ala Arg Arg Asn Arg Arg Arg Arg Trp Arg Glu Arg
35 40 45

Gln Arg Gln Ile Arg Ser Ile Ser Glu Arg Ile Leu Ser Thr Tyr Leu
50 55 60

Gly Arg Ser Ala Glu Pro Val Pro Leu Gln Leu Pro Pro Leu Glu Arg
65 70 75 80

Leu Thr Leu Asp Cys Ser Glu Asp Cys Gly Thr Ser Gly Thr Gln Gly
85 90 95

Val Gly Ser Pro Gln Ile Leu Val Glu Ser Pro Ala Val Leu Glu Ser
100 105 110

Gly Thr Lys Glu
115

<210> 59

<211> 116

<212> PRT

<213> ;

<400> 59

Met Ala Gly Arg Ser Gly Asp Ser Asp Glu Glu Leu Leu Lys Thr Val
1 5 10 15

Arg Leu Ile Lys Phe Leu Tyr Gln Ser Asn Pro Pro Pro Ser Pro Glu
20 25 30

Gly Thr Arg Gln Ala Arg Arg Asn Arg Arg Arg Arg Trp Arg Glu Arg
35 40 45

Gln Arg Gln Ile Arg Ser Ile Ser Glu Trp Ile Leu Ser Thr Tyr Leu
50 55 60

Gly Arg Pro Ala Glu Pro Val Pro Leu Gln Leu Pro Pro Leu Glu Arg
65 70 75 80

Leu Thr Leu Asp Cys Asn Glu Asp Cys Gly Thr Ser Gly Thr Gln Gly
85 90 95

Val Gly Ser Pro Gln Ile Leu Val Glu Ser Pro Thr Val Leu Glu Ser
100 105 110

Gly Thr Lys Glu
        115

<210> 60

<211> 116

<212> PRT

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade B rev pro
       tein sequence

<400> 60

Met Ala Gly Arg Ser Gly Asp Ser Asp Glu Glu Leu Leu Lys Thr Val
1   5   10   15

Arg Leu Ile Lys Phe Leu Tyr Gln Ser Asn Pro Pro Pro Ser Pro Glu
            20   25   30

Gly Thr Arg Gln Ala Arg Arg Asn Arg Arg Arg Arg Trp Arg Glu Arg
        35   40   45

Gln Arg Gln Ile Arg Ser Ile Ser Glu Trp Ile Leu Ser Thr Tyr Leu
    50   55   60

Gly Arg Pro Ala Glu Pro Val Pro Leu Gln Leu Pro Pro Leu Glu Arg
65   70   75   80

Leu Thr Leu Asp Cys Ser Glu Asp Cys Gly Thr Ser Gly Thr Gln Gly
                85   90   95

Val Gly Ser Pro Gln Ile Leu Val Glu Ser Pro Ala Val Leu Glu Ser
            100   105   110

Gly Thr Lys Glu
        115

<210> 61

<211> 101

<212> PRT

<213> Artificial sequence

<220>

<223> Most common recent ancestor reconstruction of clade B tat protein

sequence

<400> 61

Met Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser
1   5   10   15

Gln Pro Lys Thr Ala Cys Thr Asn Cys Tyr Cys Lys Lys Cys Cys Tyr
            20   25   30

His Cys Gln Val Cys Phe Ile Thr Lys Gly Leu Gly Ile Ser Tyr Gly
        35   40   45

Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln Gly Ser Gln Thr
    50   55   60

His Gln Val Ser Leu Ser Lys Gln Pro Ala Ser Gln Pro Arg Gly Asp
65   70   75   80

Pro Thr Gly Pro Lys Glu Ser Lys Lys Lys Val Glu Arg Glu Thr Glu
                85   90   95

Thr Asp Pro Val Asp
            100

<210> 62

<211> 101

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares and minimum of means reconstruction of clade B tat protein sequence

<400> 62

Met Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser
1   5   10   15

Gln Pro Lys Thr Ala Cys Thr Asn Cys Tyr Cys Lys Lys Cys Cys Phe
            20   25   30

His Cys Gln Val Cys Phe Ile Thr Lys Gly Leu Gly Ile Ser Tyr Gly
        35   40   45

Arg Lys Lys Arg Arg Gln Arg Arg Arg Ala Pro Gln Asp Ser Gln Thr
    50   55   60

His Gln Val Ser Leu Ser Lys Gln Pro Ala Ser Gln Pro Arg Gly Asp
65   70   75   80

Pro Thr Gly Pro Lys Glu Ser Lys Lys Lys Val Glu Arg Glu Thr Glu
85 90 95

Thr Asp Pro Val Asp
100

<210> 63

<211> 192

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B vif protein sequence

<400> 63

Met Glu Asn Arg Trp Gln Val Met Ile Val Trp Gln Val Asp Arg Met
1 5 10 15

Arg Ile Arg Thr Trp Lys Ser Leu Val Lys His His Met Tyr Ile Ser
20 25 30

Lys Lys Ala Lys Gly Trp Phe Tyr Arg His His Tyr Glu Ser Thr His
35 40 45

Pro Arg Ile Ser Ser Glu Val His Ile Pro Leu Gly Asp Ala Arg Leu
50 55 60

Val Ile Lys Thr Tyr Trp Gly Leu His Thr Gly Glu Arg Glu Trp His
65 70 75 80

Leu Gly Gln Gly Val Ser Ile Glu Trp Arg Lys Arg Arg Tyr Ser Thr
85 90 95

Gln Val Asp Pro Gly Leu Ala Asp Gln Leu Ile His Leu Tyr Tyr Phe
100 105 110

Asp Cys Phe Ser Glu Ser Ala Ile Arg Asn Ala Ile Leu Gly His Ile
115 120 125

Val Ser Pro Arg Cys Glu Tyr Gln Ala Gly His Asn Lys Val Gly Ser
130 135 140

Leu Gln Tyr Leu Ala Leu Thr Ala Leu Ile Thr Pro Lys Lys Ile Lys
145 150 155 160

Pro Pro Leu Pro Ser Val Arg Lys Leu Thr Glu Asp Arg Trp Asn Lys
165 170 175

Pro Gln Lys Thr Lys Gly His Arg Gly Ser His Thr Met Asn Gly His
180 185 190

<210> 64

<211> 192

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares and minimum of means reconstruction of clade B vif protein sequence

<400> 64

Met Glu Asn Arg Trp Gln Val Met Ile Val Trp Gln Val Asp Arg Met
1 5 10 15

Arg Ile Arg Thr Trp Lys Ser Leu Val Lys His His Met Tyr Ile Ser
20 25 30

Arg Lys Ala Lys Gly Trp Phe Tyr Arg His His Tyr Glu Ser Thr His
35 40 45

Pro Arg Ile Ser Ser Glu Val His Ile Pro Leu Gly Asp Ala Arg Leu
50 55 60

Val Ile Thr Thr Tyr Trp Gly Leu His Thr Gly Glu Arg Asp Trp His
65 70 75 80

Leu Gly Gln Gly Val Ser Ile Glu Trp Arg Lys Lys Arg Tyr Ser Thr
85 90 95

Gln Val Asp Pro Asp Leu Ala Asp Gln Leu Ile His Leu Tyr Tyr Phe
100 105 110

Asp Cys Phe Ser Glu Ser Ala Ile Arg Asn Ala Ile Leu Gly His Ile
115 120 125

Val Ser Pro Arg Cys Glu Tyr Gln Ala Gly His Asn Lys Val Gly Ser
130 135 140

Leu Gln Tyr Leu Ala Leu Ala Ala Leu Ile Thr Pro Lys Lys Ile Lys
145 150 155 160

Pro Pro Leu Pro Ser Val Thr Lys Leu Thr Glu Asp Arg Trp Asn Lys
165 170 175

Pro Gln Lys Thr Lys Gly His Arg Gly Ser His Thr Met Asn Gly His

180 185 190

<210> 65

<211> 96

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B vpr protein sequence

<400> 65

Met Glu Gln Ala Pro Glu Asp Gln Gly Pro Gln Arg Glu Pro Tyr Asn
1 5 10 15

Glu Trp Thr Leu Glu Leu Leu Glu Glu Leu Lys Ser Glu Ala Val Arg
20 25 30

His Phe Pro Arg Leu Trp Leu His Ser Leu Gly Gln His Ile Tyr Glu
35 40 45

Thr Tyr Gly Asp Thr Trp Ala Gly Val Glu Ala Ile Ile Arg Ile Leu
50 55 60

Gln Gln Leu Leu Phe Ile His Phe Arg Ile Gly Cys Gln His Ser Arg
65 70 75 80

Ile Gly Ile Thr Arg Gln Arg Arg Ala Arg Asn Gly Ala Ser Arg Ser
85 90 95

<210> 66

<211> 96

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares and minimum of means reconstruction of clade b vpr protein sequence

<400> 66

Met Glu Gln Ala Pro Glu Asp Gln Gly Pro Gln Arg Glu Pro Tyr Asn
1 5 10 15

Glu Trp Thr Leu Glu Leu Leu Glu Glu Leu Lys Ser Glu Ala Val Arg
20 25 30

His Phe Pro Arg Ile Trp Leu His Ser Leu Gly Gln His Ile Tyr Glu
        35   40   45

Thr Tyr Gly Asp Thr Trp Ala Gly Val Glu Ala Ile Ile Arg Ile Leu
    50   55   60

Gln Gln Leu Leu Phe Ile His Phe Arg Ile Gly Cys Arg His Ser Arg
65   70   75   80

Ile Gly Ile Thr Arg Gln Arg Arg Ala Arg Asn Gly Ala Ser Arg Ser
                85   90   95

<210> 67

<211> 96

<212> PRT

<213> Artificial sequence

<220>

<223> Minimum of means reconstructions for the clade B of vpr protein s
equence

<400> 67

Met Glu Gln Ala Pro Glu Asp Gln Gly Pro Gln Arg Glu Pro Tyr Asn
1   5   10   15

Glu Trp Thr Leu Glu Leu Leu Glu Glu Leu Lys Ser Glu Ala Val Arg
            20   25   30

His Phe Pro Arg Ile Trp Leu His Ser Leu Gly Gln His Ile Tyr Glu
        35   40   45

Thr Tyr Gly Asp Thr Trp Ala Gly Val Glu Ala Ile Ile Arg Ile Leu
    50   55   60

Gln Gln Leu Leu Phe Ile His Phe Arg Ile Gly Cys Gln His Ser Arg
65   70   75   80

Ile Gly Ile Thr Arg Gln Arg Arg Ala Arg Asn Gly Ala Ser Arg Ser
                85   90   95

<210> 68

<211> 81

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade B rev protein sequence

<400> 68

Met Gln Pro Leu Glu Ile Leu Ala Ile Val Ala Leu Val Val Ala Ala
1 5 10 15

Ile Leu Ala Ile Val Val Trp Thr Ile Val Phe Ile Glu Tyr Arg Lys
20 25 30

Ile Leu Arg Gln Arg Lys Ile Asp Arg Leu Ile Asp Arg Ile Arg Glu
35 40 45

Arg Ala Glu Asp Ser Gly Asn Glu Ser Glu Gly Asp Gln Glu Glu Leu
50 55 60

Ser Ala Leu Val Glu Met Gly His His Ala Pro Trp Asp Val Asp Asp
65 70 75 80

Leu

<210> 69

<211> 81

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares and minimum of means reconstructions for the clade B vpu protein sequences

<400> 69

Met Gln Pro Leu Gln Ile Leu Ala Ile Val Ala Leu Val Val Ala Ala
1 5 10 15

Ile Ile Ala Ile Val Val Trp Thr Ile Val Phe Ile Glu Tyr Arg Lys
20 25 30

Ile Leu Arg Gln Arg Lys Ile Asp Arg Leu Ile Asp Arg Ile Arg Glu
35 40 45

Arg Ala Glu Asp Ser Gly Asn Glu Ser Glu Gly Asp Gln Glu Glu Leu
50 55 60

Ser Ala Leu Val Glu Met Gly His His Ala Pro Trp Asp Val Asp Asp
65 70 75 80

Leu

<210> 70

<211> 1479

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade C gag protein sequence

<400> 70
atgggtgcga gagcgtcaat attaagaggg ggaaaattag atacatggga aaaaattagg 60

ttaaggccag ggggaaagaa acattatatg ataaaacacc tagtatgggc aagcagggag 120

ctggaaagat ttgcacttaa ccctggcctt ttagagacat cagaaggctg taaacaaata 180

ataaaacagc tacaaccagc tcttcagaca ggaacagagg aacttaaatc attatataac 240

acagtagcaa ctctctattg tgtacatcaa aggatagagg tacgagacac caaggaagcc 300

ttagacaaga tagaggaaga acaaaacaaa agtcagcaaa aaacacagca ggcagaagcg 360

gctgacggaa aggtcagtca aaattatcct atagtgcaga atctccaagg gcaaatggta 420

caccaggcca tatcacctag aactttgaat gcatgggtaa aagtaataga ggagaaggct 480

ttcagcccag aggtaatacc catgtttaca gcattatcag aaggagccac cccacaagat 540

ttaaacacca tgttaaatac agtgggggga catcaagcag ccatgcaaat gttaaaagat 600

accatcaatg aggaggctgc agaatgggat aggttacatc cagtgcatgc agggcctgtt 660

gcaccaggcc aaatgagaga accaagggga agtgacatag caggaactac tagtaccctt 720

caggaacaaa tagcatggat gacaagtaac ccacctatcc cagtgggaga catctataaa 780

agatggataa ttctggggtt aaataaaata gtaagaatgt atagccctgt cagcattttg 840

gacataaaac aagggccaaa ggaacccttt agagactatg tagaccggtt ctttaaaact 900

ttaagagctg aacaagctac acaagatgta aaaaattgga tgacagacac cttgttggtc 960

caaaatgcga acccagattg taagaccatt ttaagagcat taggaccagg ggctacacta 1020

gaagaaatga tgacagcatg tcagggagtg ggaggaccta gccataaagc aagagttttg 1080

gctgaggcaa tgagccaagc aaacaataca aacataatga tgcagagagg caattttaag 1140

ggccctagaa gaattgttaa atgtttcaac tgtggcaagg aaggacacat agccagaaat 1200

tgcagggccc ctaggaaaaa gggctgttgg aaatgtggaa aggaaggaca ccaaatgaaa 1260

gactgtactg agaggcaggc taattttta gggaaaattt ggccttccca caaggggagg 1320

ccagggaatt tccttcagag cagaccagag ccaacagccc caccagcaga gagcttcagg 1380

ttcgaggaga caacccccgc tccgaagcag gagccgaaag acagggaacc cttaacttcc 1440

ctcaaatcac tctttggcag cgacccсttg tctcaataa 1479

<210> 71

<211> 1479

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares center minumum of means reconstructions for clade C gag gene

<400> 71

```
atgggtgcga gagcgtcaat attaagaggc ggaaaattag atacatggga aaaaattagg      60
ttaaggccag gggaaagaa acattatatg ctaaaacacc tagtatgggc aagcagggag      120
ctggaaagat ttgcacttaa ccctggcctt ttagagacat cagaaggctg taaacaaata      180
atgaaacagc tacaaccagc tcttcagaca ggaacagagg aacttagatc attatataac      240
acagtagcaa ctctctattg tgtacatgaa aagatagagg tacgagacac caaggaagcc      300
ttagacaaga tagaggaaga acaaaacaaa agtcagcaaa aaacacagca ggcagaagcg      360
gctgacggaa aggtcagtca aaattatcct atagtgcaga atctccaagg gcaaatggta      420
caccaggcca tatcacctag aactttgaat gcatgggtaa aagtaataga ggagaaggct      480
ttcagcccag aggtaatacc catgtttaca gcattatcag aaggagccac cccacaagat      540
ttaaacacca tgttaaatac agtgggggga catcaagcag ccatgcaaat gttaaaagat      600
accatcaatg aggaggctgc agaatgggat aggttacatc cagtacatgc agggcctgtt      660
gcaccaggcc aaatgagaga accaagggga agtgacatag caggaactac tagtaccctt      720
caggaacaaa tagcatggat gacaagtaac ccacctgttc cagtgggaga catctataaa      780
agatggataa ttctggggtt aaataaaata gtaagaatgt atagccctgt cagcattttg      840
gacataaaac aagggccaaa ggaacccttt agagactatg tagaccggtt ctttaaaact      900
ttaagagctg aacaagctac acaagatgta aaaaattgga tgacagacac cttgttggtc      960
caaaatgcga acccagattg taagaccatt ttaagagcat taggaccagg ggctacatta     1020
gaagaaatga tgacagcatg tcagggagtg ggaggacctg gccacaaagc aagagtgttg     1080
gctgaggcaa tgagccaagc aaacaataca aacataatga tgcagagaag caattttaaa     1140
ggccctaaaa gaattgttaa atgtttcaac tgtggcaagg aagggcacat agccagaaat     1200
tgcagggccc ctaggaaaaa aggctgttgg aaatgtggaa aggaaggaca ccaaatgaaa     1260
gactgtactg agaggcaggc taattttta gggaaaattt ggccttccca caaggggagg     1320
ccagggaatt tccttcagag cagaccagag ccaacagccc caccagcaga gagcttcagg     1380
ttcgaggaga caacccccgc tccgaagcag gagccgaaag acagggaacc cttaacttcc     1440
ctcaaatcac tctttggcag cgaccccttg tctcaataa                            1479
```

<210> 72

<211> 1482

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means reconstructions for clade C gag gene

<400> 72

```
atgggtgcga gagcgtcaat attaagaggc ggaaaattag atacatggga aaaaattagg      60
ttaaggccag ggggaaagaa acattatatg ctaaaacacc tagtatgggc aagcagggag     120
ctggaaagat ttgcacttaa ccctggcctt ttagagacat cagaaggctg taaacaaata     180
atgaaacagc tacaaccagc tcttcagaca ggaacagagg aacttagatc attatataac     240
acagtagcaa ctctctattg tgtacatgaa aagatagagg tacgagacac caaggaagcc     300
ttagacaaga tagaggaaga acaaaacaaa agtcagcaaa aaacacagca ggcagaagcg     360
gctgctgacg gaaaggtcag tcaaaattat cctatagtgc agaatctcca agggcaaatg     420
gtacaccagg ccatatcacc tagaactttg aatgcatggg taaaagtaat agaggagaag     480
gctttcagcc cagaggtaat acccatgttt acagcattat cagaaggagc caccccacaa     540
gatttaaaca ccatgttaaa tacagtgggg ggacatcaag cagccatgca aatgttaaaa     600
gataccatca atgaggaggc tgcagaatgg gataggttac atccagtaca tgcagggcct     660
gttgcaccag gccaaatgag agaaccaagg ggaagtgaca tagcaggaac tactagtacc     720
cttcaggaac aaatagcatg gatgacaagt aacccacctg ttccagtggg agacatctat     780
aaaagatgga taattctggg gttaaataaa atagtaagaa tgtatagccc tgtcagcatt     840
ttggacataa aacaagggcc aaaggaaccc tttagagact atgtagaccg gttctttaaa     900
actttaagag ctgaacaagc tacacaagat gtaaaaaatt ggatgacaga caccttgttg     960
gtccaaaatg cgaacccaga ttgtaagacc attttaagag cattaggacc aggggctaca    1020
ttagaagaaa tgatgacagc atgtcaggga gtgggaggac ctggccacaa agcaagagtg    1080
ttggctgagg caatgagcca agcaaacaat acaaacataa tgatgcagag aagcaatttt    1140
aaaggcccta aaagaattgt taaatgtttc aactgtggca aggaagggca catagccaga    1200
aattgcaggg cccctaggaa aaaaggctgt tggaaatgtg gaaaggaagg acaccaaatg    1260
aaagactgta ctgagaggca ggctaatttt ttagggaaaa tttggccttc ccacaagggg    1320
aggccaggga atttccttca gagcagacca gagccaacag ccccaccagc agagagcttc    1380
aggttcgagg agacaacccc cgctccgaag caggagccga aagacaggga acccttaact    1440
tccctcaaat cactctttgg cagcgacccc ttgtctcaat aa                       1482
```

<210> 73

<211> 2547

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstructions of clade C env gene

<400> 73
atgagagtga tggggataca gaggaattgt caacaatggt ggatatgggg catcttaggc 60

ttttggatgt taatgatttg tagtgtggtg gggaacttgt gggtcacagt ctattatggg 120

gtacctgtgt ggaaagaagc aaaaactact ctattttgtg catcagatgc taaagcatat 180

gagagagaag tgcataatgt ctgggctaca catgcctgtg tacccacaga ccccaaccca 240

caagaaatgg ttttggaaaa tgtaacagaa aattttaaca tgtggaaaaa tgacatggtg 300

gatcagatgc atgaggatat aatcagttta tgggatcaaa gcctaaagcc atgtgtaaag 360

ttgaccccac tctgtgtcac tttaaactgt actaatgtta ataatactaa taataccaat 420

agtaccatga atggagaaat gaaaaattgc tctttcaata taaccacaga aataagagat 480

aagaagaaga aagaatatgc actttttat agacttgata tagtaccact taatgaaaat 540

aataacaata ctagtgaata tagattaata aattgtaata cctcagccat aacacaagcc 600

tgtccaaagg tctcttttga cccaattcct atacattatt gtgctccagc tggttatgcg 660

attctaaagt gtaataataa gacattcaat ggaacaggac catgcaaaaa tgtcagcaca 720

gtacaatgta cacatggaat taagccagtg gtatcaactc aactactgtt aaatggtagt 780

ctagcagaag aagagataat aattagatct gaaaatctga caaacaatgc caaaacaata 840

atagtacagc ttaatgaatc tgtagaaatt gtgtgtacaa gacccaacaa taatacaaga 900

aaaagtatga ggataggacc aggacaaaca ttctatgcaa caggagacat aataggagat 960

ataagacaag cacattgtaa cattagtgga agggaatgga ataacacttt acaacaggta 1020

gctgaaaaat taagaaaaca cttccctaat aaaacaataa aatttgcacc atcctcagga 1080

ggggacctag aaattacaac acatagcttt aattgtagag gagaattttt ctattgcaat 1140

acatcaaaac tgtttaatag tacatacaat agtacaaata gtacaaattc aaccatcaca 1200

ctcccatgca gaataaaaca aattataaac atgtggcagg gggtaggaca agcaatgtat 1260

gcccctccca ttgcaggaaa cataacatgt aaatcaaata tcacaggact actattgaca 1320

cgtgatggag gaaaaaatga aactaatgaa actgagacat tcagacctgg aggaggagat 1380

atgagggaca attggagaag tgaattatat aaatataaag tagtagaaat taaaccatta 1440

ggagtagcac ccactaaggc aaaaaggaga gtggtggaga gagaaaaaag agcagtggga 1500

ctaggagctg tgttccttgg gttcttggga gcagcaggaa gcactatggg cgcagcgtca 1560

ataacgctga cggtacaggc cagacaatta ttgtctggta tagtgcaaca gcaaagcaat 1620

ttgctgaggg ctatagaggc gcaacagcat atgttgcaac tcacagtctg gggcattaag 1680

cagctccagg caagagtcct ggctatggaa agatacctaa aggatcaaca gctcctaggg 1740

atttggggct gctctggaaa actcatctgc accactgctg tgccttggaa ctctagttgg 1800

agtaataaat ctcaagatga tatttgggat aacatgacct ggatggagtg ggatagagaa 1860

attaacaatt acacagacac aatatacagg ttgcttgaag aatcgcaaaa ccagcaggaa 1920

aaaaatgaac aagatttatt ggcattggac agttgggaaa atctgtggaa ttggtttgac 1980

atatcaaatt ggctgtggta tataaaaata ttcataatga tagtaggagg cttgataggt 2040

ttaagaataa tttttgctgt gctttctata gtaaatagag ttaggcaggg atactcacct 2100

ttgtcgtttc agacccttac cccaaacccg aggggacccg acaggctcga aagaatcgaa 2160

gaagaaggtg gagagcaaga cagagacaga tccattcgat tagtgagcgg attcttagca 2220

cttgcctggg acgacctgcg gagcctgtgc ctcttcagct accaccgctt gagagacttc 2280

atcttgattg cagcgaggac tgtggaactt ctgggacgca gcagtctcag ggactacag 2340

agggggtggg aagcccttaa atatctggga agtcttgtgc agtattgggg tcaggagcta 2400

aaaaagagtg ctattagtct gcttgatacc atagcaatag cagtagctga agggacagat 2460

aggattatag aagtagtaca aagagcttgt agagctatcc tcaacatacc tagaagaata 2520

agacagggct ttgaagcagc tttgcaa 2547

<210> 74

<211> 2547

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares and minimum of means reconstructions for clade C en
v gene

<400> 74
atgagagtga gggggatact gaggaattgt caacaatggt ggatatgggg catcttaggc 60

ttttggatgt taatgatttg taatgtggtg gggaacttgt gggtcacagt ctattatggg 120

gtacctgtgt ggaaagaagc aaaaactact ctattctgtg catcagatgc taaagcatat 180

gagaaagaag tgcataatgt ctgggctaca catgcctgtg tacccacaga ccccaaccca 240

caagaaatgg ttttggaaaa tgtaacagaa aattttaaca tgtggaaaaa tgacatggtg 300

gatcagatgc atgaggatat aatcagttta tgggatcaaa gcctaaagcc atgtgtaaag 360

ttgaccccac tctgtgtcac tttaaattgt agtaatgtta atgctaccaa tactaccaat 420

aataccatga agggagaaat aaaaaattgc tctttcaatg caaccacaga aataagagat 480

aagaaacaga aagtgtatgc actttttat agacttgata tagtaccact taatgagaat 540

aatagcaatt ctagtgagta tagattaata aattgtaata cctcagccat aacacaagcc 600
tgtccaaagg tctcttttga cccaattcct atacattatt gtgctccagc tggttatgcg 660
attctaaagt gtaataataa gacattcaat ggaacaggac catgcaataa tgtcagcaca 720
gtacaatgta cacatggaat taagccagtg gtatcaactc aactactgtt aaatggtagc 780
ctagcagaag aagagataat aattagatct gaaaatctga caaacaatgt caaaacaata 840
atagtacatc ttaatgaatc tgtagaaatt gtgtgtacaa gacccaacaa taatacaaga 900
aaaagtataa ggataggacc aggacaaaca ttctatgcaa caggagacat aataggagac 960
ataagacaag cacattgtaa cattagtgaa gaggaatgga ataaaacttt acaaagggta 1020
ggtaaaaaat tagaagaaca cttccctaat aaaacaataa aatttgaacc atcctcagga 1080
ggggacctag aaattacaac acatagcttt aattgtagag gagaattttt ctattgcaat 1140
acatcaaaac tgtttaatag tacatacaat ggtacaaata gtacaaatac aaccatcaca 1200
ctcccatgca gaataaaaca aattataaac atgtggcagg aggtaggacg agcaatgtat 1260
gcccctccca ttgcaggaaa cataacatgt aaatcaaata tcacaggact actattggta 1320
cgtgatggag gaaaaaataa cacaaataac acagagatat tcagacctgg aggaggagat 1380
atgagggaca attggagaag tgaattatat aaatataaag tggtagaaat taagccattg 1440
ggaatagcac ccactaaggc aaaaaggaga gtggtggaga gagaaaaaag agcagtggga 1500
ataggagctg tgttccttgg gttcttggga gcagcaggaa gcactatggg cgcggcgtca 1560
ataacgctga cggtacaggc cagacaattg ttgtctggta tagtgcaaca gcaaagcaat 1620
ttgctgaggg ctatagaggc gcaacagcat atgttgcaac tcacggtctg gggcattaag 1680
cagctccaga caagagtcct ggctatagaa agatacctaa aggatcaaca gctcctaggg 1740
atttggggct gctctggaaa actcatctgc accactgctg tgccttggaa ctctagttgg 1800
agtaataaat ctcaagaaga tatttgggat aacatgacct ggatgcagtg ggatagagaa 1860
attagtaatt acacagacac aatatacagg ttgcttgaag actcgcaaaa ccagcaggaa 1920
caaaatgaaa aagatttact agcattggac agttggaaaa atctgtggaa ttggtttgac 1980
ataacaaatt ggctgtggta tataaaaata ttcataatga tagtaggagg cttgataggt 2040
ttaagaataa tttttgctgt gctttctata gtgaatagag ttaggcaggg atactcacct 2100
ttgtcgtttc agacccttac ccaaacccg aggggacccg acaggctcgg aagaatcgaa 2160
gaagaaggtg gagagcaaga cagagacaga tccattcgat tagtgagcgg attcttagca 2220
cttgcctggg acgacctgcg gagcctgtgc ctcttcagct accaccgatt gagagacttc 2280
atattggtgg cagcgagagc ggtggaactt ctgggacgca gcagtctcag ggactacag 2340
agggggtggg aagcccttaa gtatctggga agtcttgtgc agtattgggg tctggagcta 2400
aaaaagagtg ctattagtct gcttgatacc atagcaatag cagtagctga aggaacagat 2460
aggattatag aattaataca aagaatttgt agagctatcc gcaacatacc tagaagaata 2520
agacagggct ttgaagcagc tttgcaa 2547

<210> 75

<211> 2550

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means reconstructions for clade C env gene

<400> 75
atgagagtga gggggatact gaggaattgt caacaatggt ggatatgggg catcttaggc 60

ttttggatgt taatgatttg taatgtggtg gggaacttgt gggtcacagt ctattatggg 120

gtacctgtgt ggaaagaagc aaaaactact ctattctgtg catcagatgc taaagcatat 180

gagaaagaag tgcataatgt ctgggctaca catgcctgtg tacccacaga ccccaaccca 240

caagaaatgg ttttggaaaa tgtaacagaa aattttaaca tgtggaaaaa tgacatggtg 300

gatcagatgc atgaggatat aatcagttta tgggatcaaa gcctaaagcc atgtgtaaag 360

ttgaccccac tctgtgtcac tttaaattgt agtaatgtta atactaccaa tactaccaat 420

aataccatga aggagaaat aaaaaattgc tctttcaatg taaccacaga actaagagat 480

aagaaaaaga aagagtatgc actttttat agacttgata tagtaccact taatgagaat 540

aataacaatt ctagtgagta tagattaata aattgtaata cctcagccat aacacaagcc 600

tgtccaaagg tctcttttga cccaattcct atacattatt gtgctccagc tggttatgcg 660

attctaaagt gtaataataa gacattcaat ggaacaggac catgcaataa tgtcagcaca 720

gtacaatgta cacatggaat taagccagtg gtatcaactc aactactgtt aaatggtagc 780

ctagcagaag aagagataat aattagatct gaaaatctga caaacaatgc caaaacaata 840

atagtacatc ttaatgaatc tgtagaaatt gtgtgtacaa gacccaacaa taatacaaga 900

aaaagtataa ggataggacc aggacaaaca ttctatgcaa caggagacat aataggagac 960

ataagacaag cacattgtaa cattagtgaa gaggaatgga ataaaacttt acaaagggta 1020

ggtaaaaaat tagaagaaca cttccctaat aaaacaataa aatttgaacc atcctcagga 1080

ggggacctag aaattacaac acatagcttt aattgtagag gagaattttt ctattgcaat 1140

acatcaaaac tgtttaatag tacatacaat ggtacaaata gtacaaattc aaccatcaca 1200

ctccaatgca gaataaaaca aattataaac atgtggcagg aggtaggacg agcaatgtat 1260

gcccctccca ttgcaggaaa cataacatgt aaatcaaata tcacaggact actattggta 1320

cgtgatggag gaaaaaatga cacaaatgac acagagatat tcagacctgg aggaggagat 1380

atgagggaca attggagaag tgaattatat aaatataaag tggtagaaat taagccattg 1440

ggaatagcac ccactaaggc aaaaaggaga gtggtggaga gagaaaaaag agcagtggga 1500

ataggagctg tgttccttgg gttcttggga gcagcaggaa gcactatggg cgcagcgtca 1560

ataacgctga cggtacaggc cagacaattg ttgtctggta tagtgcaaca gcaaagcaat 1620
ttgctgaggg ctatagaggc gcaacagcat atgttgcaac tcacggtctg ggcattaag 1680
cagctccaga caagagtcct ggctatagaa agatacctaa aggatcaaca gctcctaggg 1740
atttggggct gctctggaaa actcatctgc accactgctg tgccttggaa ctctagttgg 1800
agtaataaat ctcaagagga tatttgggat aacatgacct ggatgcagtg ggatagagaa 1860
attagtaatt acacagacac aatatacagg ttgcttgaag actcgcaaaa ccagcaggaa 1920
caaaatgaaa aagatttact agcattggac agttggaaaa atctgtggaa ttggtttgac 1980
ataacaaatt ggctgtggta tataaaaata ttcataatga tagtaggagg cttgataggt 2040
ttaagaataa tttttgctgt gctttctata gtgaatagag ttaggcaggg atactcacct 2100
ttgtcgtttc agacccttac cccaaacccg aggggacccg acaggctcgg aagaatcgaa 2160
gaagaaggtg gagagcaaga cagagacaga tccattcgat tagtgagcgg attcttagca 2220
cttgcctggg acgacctgcg gagcctgtgc ctcttcagct accaccgatt gagagacttc 2280
atattggtgg cagcgagagc ggtggaactt ctgggacgca gcagtctcag ggactacag 2340
agggggtggg aagcccttaa gtatctggga agtcttgtgc agtattgggg tctggagcta 2400
aaaaagagtg ctattagtct gcttgatacc atagcaatag cagtagctga aggaacagat 2460
aggattatag aattaataca aagaatttgt agagctatcc gcaacatacc tagaagaata 2520
agacagggct ttgaagcagc tttgcaataa 2550

<210> 76

<211> 618

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstructions of clade C nef gene

<400> 76
atgggggca agtggtcaaa aagcagtata gttggatggc ctgctgtaag agaaagaata 60
agacgaactg ctccagcagc agaaggagta ggagcagcgt ctcaagactt agataaacat 120
ggagcactta caagcagcaa cacagccgcc actaatgctg attgtgcctg gctggaagca 180
caagaggagg aagaagtagg ctttccagtc agacctcagg tgcctttaag accaatgact 240
tataagggag cagtcgatct cagcttcttt ttaaaagaaa agggggact ggaagggtta 300
atttactcta agaaaaggca agagatcctt gatttgtggg tctatcacac acaaggctac 360
ttccctgatt ggcaaaacta cacaccggga ccagggatca gatttccact gacctttgga 420
tggtgcttca agctagtgcc agttgaccca agggaagtag aagaggccaa tgaaggagag 480
aacaactgct tgctacaccc tatgagccag catggaatgg aggatgaaga cagagaagta 540

ttaaagtgga agtttgacag tcacctagca cgcagacaca tggcccgcga gctacatccg 600

gagtattaca aagactgc 618

<210> 77

<211> 624

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares and minimum of means reconstructions of clade C nef gene

<400> 77
atgggggca agtggtcaaa aagcagtata gttggatggc ctgctgtaag agaaagaata 60

agacgaactg agccagcagc agagggagta ggagcagcgt ctcaagactt agataaacat 120

ggagcactta caagcagcaa cacagccgcc aataatgctg attgtgcctg gctggaagca 180

caagaggagg aagaagaagt aggctttcca gtcagacctc aggtgccttt aagaccaatg 240

acttataagg gagcattcga tctcagcttc tttttaaaag aaaagggggg actggaaggg 300

ttaatttact ctaagaaaag gcaagagatc cttgatttgt gggtctatca cacacaaggc 360

tacttccctg attggcaaaa ctacacaccg ggaccagggg tcagatatcc actgaccttt 420

ggatggtgct tcaagctagt gccagttgac ccaagggaag tagaagaggc caacgaagga 480

gagaacaact gtttgctaca ccctatgagc cagcatggaa tggaggatga agacagagaa 540

gtattaaagt ggaagtttga cagtcaccta gcacgcagac acatggcccg cgagctacat 600

ccggagtatt acaaagactg ctga 624

<210> 78

<211> 3000

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstructions of clade C pol gene

<400> 78
tttttaggg aaaatttggc cttcccacaa ggggaggcca gggaatttcc ttcagagcag 60

accagagcca acagccccac cagcagagag cttcaggttc gaggagacaa cccccgctcc 120

gaagcaggag ccgaaagaca gggaaccctt aacttccctc aaatcactct ttggcagcga 180

cccccttgtct caataaaagt agggggccag ataaaggaag ctctattaga tacaggagca 240

gatgatacag tattagaaga cataaatttg ccaggaaaat ggaaaccaaa aatgataggg 300

ggaattggag gttttatcaa agtaagacag tatgatcaaa tacttataga aatttgtgga 360
aaaaaggcta taggtacagt attagtagga cctacacctg tcaacataat tggaagaaat 420
atgttgactc agcttggttg cactctaaat tttccaatta gtcctattga aactgtacca 480
gtaaaattaa agccaggaat ggatggccca aaggttaaac aatggccatt gacagaagag 540
aaaataaaag cattaacagc aatttgtgaa gaaatggaaa aggaaggaaa aattacaaaa 600
attgggcctg aaaatccata taacactcca gtatttgcca taaaaaagaa ggacagtact 660
aagtggagaa aattagtaga tttcagagaa ctcaataaaa gaactcaaga cttctgggaa 720
gttcaattag gaataccaca cccagcaggg ttaaaaaaga aaaaatcagt aacagtactg 780
gatgtggggg atgcatattt ttcagttcct ttagatgaag acttcaggaa atatactgca 840
ttcaccatac ctagtataaa caatgaaaca ccagggatta gatatcaata taatgtgctt 900
ccacagggat ggaaaggatc accagcaata ttccagagta gcatgacaaa aatcttagag 960
ccctttaggg cacaaaaccc agaaatagtt atctatcaat acatggatga cttgtatgta 1020
ggatctgact tagaaatagg gcaacataga gcaaaaatag aggagttaag agaacatcta 1080
ttgaaatggg gatttaccac accagacaag aaacatcaga aagaaccccc atttctttgg 1140
atggggtatg aactccatcc tgacaaatgg acagtacagc ctatacagct gccagaaaag 1200
gatagctgga ctgtcaatga tatacagaag ttagtgggaa aattaaactg ggcaagtcag 1260
atttacccag ggattaaagt aaggcaactg tgtaaactcc ttaggggagc caaagcacta 1320
acagacatag taccactgac tgaagaagca gaattagaat tggcagagaa cagggaaatt 1380
ctaaaagaac cagtacatgg agtatattat gacccatcaa aagacttaat agctgaaata 1440
cagaaacagg ggcatgacca atggacatat caaatttacc aagaaccatt caaaaatctg 1500
aaaacaggaa agtatgcaaa aatgaggtct gcccacacta atgatgtaaa acaattaaca 1560
gaagcagtgc aaaaaatagc catggaaagc atagtaatat ggggaaagac tcctaaattt 1620
agactaccca tccaaaaaga aacatgggag acatggtgga cagactattg gcaagccacc 1680
tggattcctg agtgggagtt tgttaatacc cctcccctag taaaattatg gtaccagcta 1740
gaaaaagaac ccatagcagg agcagaaact ttctatgtag atggggcagc taatagggaa 1800
actaaactag gaaaagcagg gtatgttact gacaaaggaa gacagaaagt tgtttctcta 1860
actgaaacaa caaatcagaa gactgaatta caagcaattc agctagcttt gcaggattca 1920
ggatcagaag taaacatagt aacagactca caatatgcat taggaatcat tcaagcacaa 1980
ccagataaga gtgaatcaga gttagtcaat caaataatag agcagttaat aaaaaaggaa 2040
aaggtctacc tgtcatgggt accagcacat aaaggaattg gaggaaatga acaagtagat 2100
aaattagtaa gttctggaat caggaaagtg ctgtttctag atggaataga taaagctcaa 2160
gaagaacatg aaaaatatca cagcaattgg agagcaatgg ctagtgagtt taatctgcca 2220
cccatagtag caaaagaaat agtagctagc tgtgataaat gtcagctaaa aggggaagcc 2280
atgcatggac aagtagactg tagtccaggg atatggcaat tagattgtac acatttagaa 2340

ggaaaagtta tcctggtagc agtccatgta gccagtggct acatagaagc agaagttatc 2400
ccagcagaaa caggacagga aacagcatac tttatattaa aattagcagg aagatggcca 2460
gtaaaagtaa tacatacaga caatggcagc aatttcacca gtgctgcagt taaggcagcc 2520
tgttggtggg caggtatcca acaggaattt ggaattccct acaatcccca aagtcaggga 2580
gtagtagaat ccatgaataa agaattaaag aaaatcatag ggcaggtaag agatcaagct 2640
gagcacctta agacagcagt acaaatggca gtattcattc acaattttaa aagaaaaggg 2700
gggattgggg ggtacagtgc aggggaaaga ataatagaca taatagcaac agacatacaa 2760
actaaagaat tacaaaaaca aattataaaa attcaaaatt ttcgggttta ttacagagac 2820
agcagagacc ctgtttggaa aggaccagcc aaactactct ggaaaggtga aggggcagta 2880
gtaatacaag acaatagtga cataaaggta gtaccaagga ggaaagcaaa gatcattagg 2940
gattatggaa aacagatggc aggtgctgat tgtgtggcag gtagacagga tgaagattag 3000

<210> 79

<211> 2999

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares reconstruction for clade c pol gene

<400> 79
tttttagggg aaaatttggc cttcccacaa ggggaggcca gggaatttcc ttcagagcag 60
accagagcca acagccccac cagcagagag cttcaggttc gaggagacaa cccccgctcc 120
gaagcaggag ccgaaagaca gggaaccctt aacttccctc aaatcactct ttggcagcga 180
ccccttgtct caataaaagt agggggccag ataaaggagg ctctcttaga cacaggagca 240
gatgatacag tattagaaga aataaatttg ccaggaaaat ggaaaccaaa aatgatagga 300
ggaattggag gttttatcaa agtaagacag tatgatcaaa tacttataga aatttgtgga 360
aaaaaggcta taggtacagt attagtagga cctacacctg tcaacataat tggaagaaat 420
atgttgactc agcttggatg cacactaaat ttttccaatta gtcccattga aactgtacca 480
gtaaaattaa agccaggaat ggatggccca aaggttaaac aatggccatt gacagaagag 540
aaaataaaac attaacagca atttgtgaag aaatggagaa ggaaggaaaa attacaaaaa 600
ttgggcctga aaatccatat aacactccag tatttgccat aaaaaagaag gacagtacta 660
agtggagaaa attagtagat ttcagggaac tcaataaaag aactcaagac ttttgggaag 720
ttcaattagg aataccacac ccagcagggt taaaaaagaa aaaatcagtg acagtactgg 780
atgtgggga tgcatatttt tcagttcctt tagatgaagg cttcaggaaa tatactgcat 840
tcaccatacc tagtataaac aatgaaacac cagggattag atatcaatat aatgtgcttc 900

cacagggatg gaaaggatca ccagcaatat tccagagtag catgacaaaa atcttagagc 960
cctttagggc acaaaatcca gaaatagtca tctatcaata tatggatgac ttgtatgtag 1020
gatctgactt agaaataggg caacatagag caaaaataga ggagttaaga gaacatctat 1080
taaagtgggg atttaccaca ccagacaaga aacatcagaa agaaccccca tttctttgga 1140
tggggtatga actccatcct gacaaatgga cagtacagcc tatacagctg ccagaaaagg 1200
atagctggac tgtcaatgat atacagaagt tagtgggaaa attaaactgg gcaagtcaga 1260
tttacccagg gattaaagta aggcaacttt gtaaactcct taggggggcc aaagcactaa 1320
cagacatagt accactaact gaagaagcag aattagaatt ggcagagaac agggaaattc 1380
taaaagaacc agtacatgga gtatattatg acccatcaaa agacttgata gctgaaatac 1440
agaaacaggg gcatgaccaa tggacatatc aaatttacca agaaccattc aaaaatctga 1500
aaacagggaa gtatgcaaaa atgaggactg cccacactaa tgatgtaaaa cagttaacag 1560
aggcagtgca aaaaatagcc atggaaagca tagtaatatg gggaaagact cctaaattta 1620
gactacccat ccaaaaagaa acatgggaga catggtggac agactattgg caagccacct 1680
ggattcctga gtgggagttt gttaataccc ctcccctagt aaaattatgg taccagctgg 1740
agaaagaacc catagcagga gcagaaactt tctatgtaga tggagcagct aatagggaaa 1800
ctaaaatagg aaaagcaggg tatgttactg acagaggaag gcagaaaatt gtttctctaa 1860
ctgaaacaac aaatcagaag actgaattac aagcaattca gctagctttg caagattcag 1920
gatcagaagt aaacatagta acagactcac agtatgcatt aggaatcatt caagcacaac 1980
cagataagag tgaatcagag ttagtcaacc aaataataga acaattaata aaaaaggaaa 2040
gggtctacct gtcatgggta ccagcacata aaggaattgg aggaaatgaa caagtagata 2100
aattagtaag tagtggaatc aggaaagtgc tgtttctaga tggaatagat aaggctcaag 2160
aagagcatga aaagtatcac agcaattgga gagcaatggc tagtgagttt aatctgccac 2220
ccatagtagc aaaagaaata gtagctagct gtgataaatg tcagctaaaa ggggaagcca 2280
tacatggaca agtagactgt agtccaggga tatggcaatt agattgtaca catttagaag 2340
gaaaaatcat cctggtagca gtccatgtag ccagtggcta catagaagca gaggttatcc 2400
cagcagaaac aggacaagaa acagcatact ttatactaaa attagcagga agatggccag 2460
tcaaagtaat acatacagac aatggcagta atttcaccag tgctgcagtt aaggcagcct 2520
gttggtgggc aggtatccaa caggaatttg gaattcccta caatccccaa agtcagggag 2580
tagtagaatc catgaataaa gaattaaaga aaatcatagg gcaggtaaga gatcaagctg 2640
agcaccttaa gacagcagta caaatggcag tattcattca caattttaaa agaaaagggg 2700
ggattggggg gtacagtgca ggggaaagaa taatagacat aatagcaaca gacatacaaa 2760
ctaaagaatt acaaaaacaa attataaaaa ttcaaaattt tcgggtttat tacagagaca 2820
gcagagaccc tatttggaaa ggaccagcca aactactctg gaaaggtgaa ggggcagtag 2880
taatacaaga taatagtgac ataaaggtag taccaaggag gaaagcaaaa atcattaagg 2940

actatggaaa acagatggca ggtgctgatt gtgtggcagg tagacaggat gaagattag 2999

<210> 80

<211> 3000

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means reconstructions of clade c pol gene

<400> 80

```
tttttaggg aaaatttggc cttcccacaa ggggaggcca gggaatttcc ttcagagcag      60
accagagcca acagccccac cagcagagag cttcaggttc gaggagacaa ccccctctcc    120
gaagcaggag ccgaaagaca gggaaccctt aacttccctc aaatcactct ttggcagcga    180
cccctgtca caataaaagt agggggacag ctaaaggagg ctctcttaga cacaggagca    240
gatgatacag tattagaaga aataaatttg ccaggaaaat ggaaaccaaa aatgatagga    300
ggaattggag gttttatcaa agtaagacag tatgatcaaa tacttataga aatttgtgga    360
aaaaaggcta taggtacagt actagtagga cctacacctg tcaacataat tggaagaaat    420
atgttgactc agcttggatg cacactaaat tttccaatta gtcccattga aactgtacca    480
gtaaaattaa agccaggaat ggatggccca aaggtcaaac aatggccatt gacagaagag    540
aaaataaaag cattaacagc aatttgtgaa gaaatggaga aggaaggaaa aattacaaaa    600
attgggcctg aaaatccata taacactcca gtatttgcca taaaaaagaa ggacagtact    660
aagtggagaa aattagtaga tttcagggaa ctcaataaaa gaactcaaga cttttgggaa    720
gttcaattag ggataccaca cccagcaggg ttaaaaaaga aaaaatcagt gacagtactg    780
gatgtggggg atgcatattt ttcagttcct ttagatgaag gcttcaggaa atatactgca    840
ttcaccatac ctagtataaa caatgaaaca ccagggatta gatatcaata taatgtgctt    900
ccacagggat ggaaaggatc accagcaata ttccagagta gcatgacaaa aatcttagag    960
ccctttaggg cacaaaatcc agaaatagtt atctatcaat atatggatga cttgtatgta   1020
ggatctgact tagaaatagg gcaacataga gcaaaaatag aggagttaag agaacatcta   1080
ttgaagtggg gatttaccac accagacaag aaacatcaga aagaaccccc atttctttgg   1140
atggggtatg aactccatcc tgacaaatgg acagtacagc ctatacagct gccagaaaag   1200
gatagctgga ctgtcaatga tatacagaag ttagtgggaa aattaaactg ggcaagtcag   1260
atttacccag ggattaaagt aaggcaactg tgtaaactcc ttaggggagc caaagcacta   1320
acagacatag taccactaac tgaagaagca gaattagaat tggcagagaa cagggaaatt   1380
ctaaaagaac cagtacatgg agtatattat gacccatcaa aagacttaat agctgaaata   1440
cagaaacagg ggcatgacca atggacatat caaatttacc aagaaccatt caaaaatctg   1500
```

aaaacaggga agtatgcaaa aatgaggact gcccacacta atgatgtaaa acagttaaca 1560
gaggcagtgc aaaaaatagc catggaaagc atagtaatat ggggaaagac tcctaaattt 1620
agattaccca tccagaaaga aacatgggag gcatggtgga cagactattg gcaagccacc 1680
tggattcctg agtgggagtt tgttaatacc cctcccctag taaaattatg gtaccagctg 1740
gagaaagaac ccatagcagg agcagaaact ttctatgtag atggagcagc taatagggaa 1800
actaaaatag gaaaagcagg gtatgttact gacagaggaa ggcagaaaat tgtttctcta 1860
actgaaacaa caaatcagaa gactgaatta caagcaattc agctagcttt gcaggattca 1920
ggatcagaag taaacatagt aacagactca cagtatgcat taggaatcat tcaagcacaa 1980
ccagataaga gtgaatcaga gttagtcaat caaataatag aacagttaat aaaaaaggaa 2040
agggtctacc tgtcatgggt accagcacat aaaggaattg gaggaaatga acaagtagat 2100
aaattagtaa gtagtggaat caggaaagtg ctgtttctag atggaataga taaggctcaa 2160
gaagagcatg aaaaatatca cagcaattgg agagcaatgg ctagtgagtt taatctgcca 2220
cccatagtag caaaagaaat agtagctagc tgtgataaat gtcagctaaa aggggaagcc 2280
atacatggac aagtagactg tagtccaggg atatggcaat tagattgtac acatttagaa 2340
ggaaaaatca tcctggtagc agtccatgta gccagtggct acatagaagc agaggttatc 2400
ccagcagaaa caggacaaga aacagcatac tttatactaa aattagcagg aagatggcca 2460
gtcaaagtaa tacatacaga caatggcagt aatttcacca gtgctgcagt taaagcagcc 2520
tgttggtggg caggtatcca acaggaattt ggaattccct acaatcccca aagtcaggga 2580
gtagtagaat ccatgaataa agaattaaag aaaatcatag ggcaggtaag agatcaagct 2640
gagcacctta agacagcagt acaaatggca gtattcattc acaattttaa aagaaaaggg 2700
gggattgggg ggtacagtgc aggggaaaga ataatagaca taatagcaac agacatacaa 2760
actaaagaat tacaaaaaca aattataaaa attcaaaatt ttcgggttta ttacagagac 2820
agcagagacc ctatttggaa aggaccagcc aaactactct ggaaaggtga aggggcagta 2880
gtaatacaag ataacagtga cataaaggta gtaccaagga ggaaagcaaa aatcattaag 2940
gactatggaa aacagatggc aggtgctgat tgtgtggcag gtagacagga tgaagattag 3000

<210> 81

<211> 381

<212> DNA

<213> Artificial sequence

<220>

<223> Most common recent ancestor reconstructions of clade c rev gene

<400> 81
atggcaggaa gaagcggaga cagcgacgaa gcgctcctcc aagcagtgag gatcatcaaa 60

atcctatatc aaagcaaccc ttaccccaaa cccgagggga cccgacaggc tcgaaggaat 120

cgaagaagaa ggtggagagc aagacagaga cagatccatt cgattagtga gcggattctt 180

agcacttgcc tgggacgacc tgcggagcct gtgcctcttc agctaccacc gcttgagaga 240

cttcatcttg attgcagcga ggactgtgga acttctggga cgcagcagtc tcaggggact 300

acagaggggg tgggaagccc ttaaatatct gggaagcctt gtgcagtatt ggggtcagga 360

gctaaaaaag agtgctatta g 381

<210> 82

<211> 381

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares of reconstructions of clade c rev gene

<400> 82
atggcaggaa gaagcggaga cagcgacgaa gcgctcctcc aagcagtgag gatcatcaaa 60

atcttatatc aaagcaaccc ttaccccaaa cccgagggga cccgacaggc tcggaagaat 120

cgaagaagaa ggtggagagc aagacagaga cagatccatt cgattagtga gcggattctt 180

agcacttgcc tgggacgacc tgcggagcct gtgcctcttc agctaccacc gattgagaga 240

cttcatattg gtgacagcga gagcagtgga acttctggga cgcagcagtc tcaggggact 300

acagaggggg tgggaagccc ttaagtatct gggaagtctt gtgcagtatt ggggtctgga 360

actaaaaaag agtgctatta g 381

<210> 83

<211> 381

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means reconstruction of clade C rev gene

<400> 83
atggcaggaa gaagcggaga cagcgacgaa gcgctcctcc aagcagtgag gatcatcaaa 60

atcctatatc aaagcaaccc ttaccccaaa cccgagggga cccgacaggc tcggaagaat 120

cgaagaagaa ggtggagagc aagacagaga cagatccatt cgattagtga gcggattctt 180

agcacttgcc tgggacgacc tgcggagcct gtgcctttc agctaccacc gattgagaga 240

cttcatattg gtgacagcga gagcagtgga acttctggga cgcagcagtc tcaggggact 300

acagaggggg tgggaagccc ttaagtatct gggaagtctt gtgcagtatt ggggtctgga 360

actaaaaaag agtgctatta g 381

<210> 84

<211> 306

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade C tat gene se
quence

<400> 84
atggagccag tagatcctaa cctagagccc tggaaccatc caggaagtca gcctaaaact 60

gcttgtaata aatgttattg taaaaaatgt agctatcatt gtctagtttg ctttctgaca 120

aaaggcttag gcatttccta tggcaggaag aagcggagac agcgacgaag agctcctcca 180

agcagtgagg atcatcaaaa tcctatatca aagcaaccct tatcccaaac ccgaggggac 240

ccgacaggct cggaggaatc gaagaagaag gtggagagca agacagagac agatccgtgc 300

gattag 306

<210> 85

<211> 306

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares of reconstruction of clade C tat gene sequence

<400> 85
atggagccag tagatcctaa cctagagccc tggaaccatc caggaagtca gcctaaaact 60

ccttgtaata agtgttattg taaacactgt agctatcatt gtctagtttg ctttcagaca 120

aaaggcttag gcatttccta tggcaggaag aagcggagac agcgacgaag cgctcctcca 180

agcagtgagg atcatcaaaa tcctatatca aagcaaccct taccccaaac ccgaggggac 240

ccgacaggct cggaagaatc gaagaagaag gtggagagca agacagagac agatccattc 300

gattag 306

<210> 86

<211> 306

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means reconstructions of clade C tat gene sequence

<400> 86
atggagccag tagatcctaa cctagagccc tggaaccatc caggaagtca gcctaaaact 60

ccttgtaata agtgttattg taaacactgt agctatcatt gtctagtttg ctttcagaca 120

aaaggcttag gcatttccta tggcaggaag aagcggagac agcgacgaag cgctcctcca 180

agcagtgagg atcatcaaaa tcctatatca aagcaacccct taccccaaac ccgaggggac 240

ccgacaggct cggaggaatc gaagaagaag gtggagagca agacagagac agatccattc 300

gattag 306

<210> 87

<211> 579

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstructions of clade C vif gene s equence

<400> 87
atggaaaaca gatggcaggt gctgattgtg tggcaggtag acaggatgaa gattagaaca 60

tggaatagtt tagtaaaaca ccatatgtat gtttcaagga gagctaaagg atggttttat 120

agacatcact atgaaagcag acatccaaaa ataagttcag aagtacacat cccattaggg 180

gatgctagat tagtaataaa aacatattgg ggtttgcata caggagaaag agattggcat 240

ttgggtcatg gagtctccat agaatggaga ctgagaagat atagcacaca agtagaccct 300

ggcctggcag accaactaat tcatatgcat tattttgatt gttttgcaga ctctgccata 360

aggaaagcca tattaggaca tatagttagc cctaggtgtg actatcaagc aggacataac 420

aaggtaggat ctctacaata cttggcactg acagcattaa taaaaccaaa aaagataaag 480

ccacctctgc ctagtgttaa gaaattagta gaggatagat ggaacaagcc ccagaagacc 540

aggggccaca gagggagcca tacaatgaat ggacactag 579

<210> 88

<211> 579

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares of reconstruction of clade C vif gene sequence

<400> 88
atggaaaaca gatggcaggt gctgattgtg tggcaggtag acaggatgaa gattagaaca 60

tggaatagtt tagtaaagca ccatatgtat gtttcaagga gagctaatgg atggttttac 120

agacatcatt atgaaagcag acatccaaaa gtaagttcag aagtacacat cccattaggg 180

gatgctagat tagtaataaa aacatattgg ggtttgcaaa caggagaaag agattggcat 240

ttgggtcatg gagtctccat agaatggaga ttgagaagat atagcacaca agtagaccct 300

ggcctggcag accagctaat tcatatgcat tattttgatt gttttgcaga ctctgccata 360

agaaaagcca tattaggaca catagttatt cctaggtgtg actatcaagc aggacataat 420

aaggtaggat ctctacaata cttggcactg acagcattga taaaaccaaa aaagataaag 480

ccacctctgc ctagtgttag gaaattagta gaggatagat ggaacaagcc ccagaagacc 540

aggggccgca gagggaacca tacaatgaat ggacactag 579

<210> 89

<211> 579

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade C vif gen
e sequence

<400> 89
atggaaaaca gatggcaggt gctgattgtg tggcaggtag acaggatgaa gattagaaca 60

tggaatagtt tagtaaagca ccatatgtat gtttcaagga gagctaatgg atggttttac 120

agacatcatt atgaaagcag acatccaaaa gtaagttcag aagtacacat cccattaggg 180

gatgctagat tagtaataaa aacatattgg ggtttgcata caggagaaag agattggcat 240

ttgggtcatg gagtctccat agaatggaga ttgagaagat atagcacaca agtagaccct 300

ggcctggcag accagctaat tcatatgcat tattttgatt gttttgcaga ctctgccata 360

agaaaagcca tattaggaca catagttatt cctaggtgtg actatcaagc aggacataat 420

aaggtaggat ctctacaata cttggcactg acagcattga taaaaccaaa aaagataaag 480

ccacctctgc ctagtgttag gaaattagta gaggatagat ggaacaagcc ccagaagacc 540

aggggccgca gagggaacca tacaatgaat ggacactag 579

<210> 90

<211> 291

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstructions for clade C vpr gene sequence

<400> 90
atggaacaag ccccagaaga ccaggggcca cagagggagc catacaatga atggacacta 60

gagcttttag aggaacttaa gcaggaagct gtcagacatt ttcctagacc atggctccat 120

agcttaggac aacatatcta tgaaacctat ggggatactt gggcgggagt tgaagctata 180

ataagaattc tgcaacaact actgtttatt catttcagaa ttgggtgcca acatagcaga 240

ataggcatta ttcgacagag aagagcaaga aatggagcca gtagatccta a 291

<210> 91

<211> 291

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade C vpr gene s equence

<400> 91
atggaacaag ccccagaaga ccaggggccg cagagggaac catacaatga atggacacta 60

gagattttag aggaactcaa gcaggaagct gtcagacact ttcctagacc atggctccat 120

agcttaggac aatatatcta tgaaacctat ggggatactt ggacaggagt cgaagctcta 180

ataagaatac tgcaacaact actgtttatt catttcagaa ttgggtgcca gcatagcaga 240

ataggcattt tgcgacagag aagagcaaga aatggagcca gtagatccta a 291

<210> 92

<211> 291

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade C vpr gen e sequence

<400> 92
atggaacaag ccccagaaga ccaggggccg cagagggaac catacaatga atggacacta 60

gagcttttag aggaactcaa gcaggaagct gtcagacact ttcctagacc atggctccat 120

agcttaggac aacatatcta tgaaacctat ggggatactt ggacgggagt tgaagctcta 180

ataagaattc tgcaacaact actgtttatt catttcagaa ttgggtgcca gcatagcaga 240

ataggcatta tgcgacagag aagagcaaga aatggagcca gtagatccta a 291

<210> 93

<211> 261

<212> DNA

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade C vpu gene se
quence

<400> 93
atgttagatt taatagcaag agtagattat agattaggag taggagcatt gatagtagca 60

ctaatcatag caatagttgt gtggaccata gtatatatag aatataggaa attggtaaga 120

caaagaaaaa tagactggtt aattaaaaga attagggaaa gagcagaaga cagtggcaat 180

gagagtgatg gggatacaga ggaattgtca acactggtgg atatggggca tcttaggctt 240

ttggatgtta atgatttgta a 261

<210> 94

<211> 261

<212> DNA

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade C vpu gene s
equence

<400> 94
atgttagatt tactagcaag agtagattat agattaggag taggagcatt gatagtagca 60

ctaatcatag caatagttgt gtggaccata gtatatatag aatataggaa attgttaaga 120

caaagaaaaa tagactggtt aattaaaaga attagggaaa gagcagaaga cagtggcaat 180

gagagtgagg gggatactga ggaattgtca acaatggtgg atatggggca tcttaggctt 240

ttggatgtta atgatttgta a 261

<210> 95

<211> 261

<212> DNA

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade C vpu gen
e sequence

<400> 95
atgttagatt tactagcaag agtagattat agattaggag taggagcatt gatagtagca 60

ctaatcatag caatagttgt gtggaccata gtatatatag aatataggaa attgttaaga 120

caaagaaaaa tagactggtt aattaaaaga attagggaaa gagcagaaga cagtggcaat 180

gagagtgagg gggatactga ggaattatca acaatggtgg atatggggca tcttaggctt 240

ttggatgtta atgatttgta a 261

<210> 96

<211> 492

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade C gag protein
sequence

<400> 96

Met Gly Ala Arg Ala Ser Ile Leu Arg Gly Gly Lys Leu Asp Thr Trp
1 5 10 15

Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys His Tyr Met Ile Lys
20 25 30

His Leu Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Leu Asn Pro
35 40 45

Gly Leu Leu Glu Thr Ser Glu Gly Cys Lys Gln Ile Ile Lys Gln Leu
50 55 60

Gln Pro Ala Leu Gln Thr Gly Thr Glu Glu Leu Lys Ser Leu Tyr Asn
65 70 75 80

Thr Val Ala Thr Leu Tyr Cys Val His Gln Arg Ile Glu Val Arg Asp
85 90 95

Thr Lys Glu Ala Leu Asp Lys Ile Glu Glu Glu Gln Asn Lys Ser Gln
100 105 110

```
Gln Lys Thr Gln Gln Ala Glu Ala Ala Asp Gly Lys Val Ser Gln Asn
        115                 120                 125

Tyr Pro Ile Val Gln Asn Leu Gln Gly Gln Met Val His Gln Ala Ile
    130                 135                 140

Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Ile Glu Glu Lys Ala
145                 150                 155                 160

Phe Ser Pro Glu Val Ile Pro Met Phe Thr Ala Leu Ser Glu Gly Ala
                165                 170                 175

Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly Gly His Gln
            180                 185                 190

Ala Ala Met Gln Met Leu Lys Asp Thr Ile Asn Glu Glu Ala Ala Glu
        195                 200                 205

Trp Asp Arg Leu His Pro Val His Ala Gly Pro Val Ala Pro Gly Gln
    210                 215                 220

Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr Ser Thr Leu
225                 230                 235                 240

Gln Glu Gln Ile Ala Trp Met Thr Ser Asn Pro Pro Ile Pro Val Gly
                245                 250                 255

Asp Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys Ile Val Arg
            260                 265                 270

Met Tyr Ser Pro Val Ser Ile Leu Asp Ile Lys Gln Gly Pro Lys Glu
        275                 280                 285

Pro Phe Arg Asp Tyr Val Asp Arg Phe Phe Lys Thr Leu Arg Ala Glu
    290                 295                 300

Gln Ala Thr Gln Asp Val Lys Asn Trp Met Thr Asp Thr Leu Leu Val
305                 310                 315                 320

Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Arg Ala Leu Gly Pro
                325                 330                 335

Gly Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly Val Gly Gly
            340                 345                 350

Pro Ser His Lys Ala Arg Val Leu Ala Glu Ala Met Ser Gln Ala Asn
        355                 360                 365

Asn Thr Asn Ile Met Met Gln Arg Gly Asn Phe Lys Gly Pro Arg Arg
    370                 375                 380
```

Ile Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His Ile Ala Arg Asn
385   390   395   400

Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys Gly Lys Glu Gly
                405   410   415

His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn Phe Leu Gly Lys
            420   425   430

Ile Trp Pro Ser His Lys Gly Arg Pro Gly Asn Phe Leu Gln Ser Arg
        435   440   445

Pro Glu Pro Thr Ala Pro Pro Ala Glu Ser Phe Arg Phe Glu Glu Thr
    450   455   460

Thr Pro Ala Pro Lys Gln Glu Pro Lys Asp Arg Glu Pro Leu Thr Ser
465   470   475   480

Leu Lys Ser Leu Phe Gly Ser Asp Pro Leu Ser Gln
                485   490

<210> 97

<211> 492

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade C gag protei
       n sequence

<400> 97

Met Gly Ala Arg Ala Ser Ile Leu Arg Gly Gly Lys Leu Asp Thr Trp
1   5   10   15

Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys His Tyr Met Leu Lys
            20   25   30

His Leu Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Leu Asn Pro
        35   40   45

Gly Leu Leu Glu Thr Ser Glu Gly Cys Lys Gln Ile Met Lys Gln Leu
    50   55   60

Gln Pro Ala Leu Gln Thr Gly Thr Glu Glu Leu Arg Ser Leu Tyr Asn
65   70   75   80

Thr Val Ala Thr Leu Tyr Cys Val His Glu Lys Ile Glu Val Arg Asp
                85   90   95

Thr Lys Glu Ala Leu Asp Lys Ile Glu Glu Glu Gln Asn Lys Ser Gln
100 105 110

Gln Lys Thr Gln Gln Ala Glu Ala Ala Asp Gly Lys Val Ser Gln Asn
115 120 125

Tyr Pro Ile Val Gln Asn Leu Gln Gly Gln Met Val His Gln Ala Ile
130 135 140

Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Ile Glu Glu Lys Ala
145 150 155 160

Phe Ser Pro Glu Val Ile Pro Met Phe Thr Ala Leu Ser Glu Gly Ala
165 170 175

Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly Gly His Gln
180 185 190

Ala Ala Met Gln Met Leu Lys Asp Thr Ile Asn Glu Glu Ala Ala Glu
195 200 205

Trp Asp Arg Leu His Pro Val His Ala Gly Pro Val Ala Pro Gly Gln
210 215 220

Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr Ser Thr Leu
225 230 235 240

Gln Glu Gln Ile Ala Trp Met Thr Ser Asn Pro Pro Val Pro Val Gly
245 250 255

Asp Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys Ile Val Arg
260 265 270

Met Tyr Ser Pro Val Ser Ile Leu Asp Ile Lys Gln Gly Pro Lys Glu
275 280 285

Pro Phe Arg Asp Tyr Val Asp Arg Phe Phe Lys Thr Leu Arg Ala Glu
290 295 300

Gln Ala Thr Gln Asp Val Lys Asn Trp Met Thr Asp Thr Leu Leu Val
305 310 315 320

Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Arg Ala Leu Gly Pro
325 330 335

Gly Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly Val Gly Gly
340 345 350

Pro Gly His Lys Ala Arg Val Leu Ala Glu Ala Met Ser Gln Ala Asn
355 360 365

Asn Thr Asn Ile Met Met Gln Arg Ser Asn Phe Lys Gly Pro Lys Arg
370 375 380

Ile Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His Ile Ala Arg Asn
385 390 395 400

Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys Gly Lys Glu Gly
405 410 415

His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn Phe Leu Gly Lys
420 425 430

Ile Trp Pro Ser His Lys Gly Arg Pro Gly Asn Phe Leu Gln Ser Arg
435 440 445

Pro Glu Pro Thr Ala Pro Pro Ala Glu Ser Phe Arg Phe Glu Glu Thr
450 455 460

Thr Pro Ala Pro Lys Gln Glu Pro Lys Asp Arg Glu Pro Leu Thr Ser
465 470 475 480

Leu Lys Ser Leu Phe Gly Ser Asp Pro Leu Ser Gln
485 490

<210> 98

<211> 493

<212> PRT

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade C gag pro tein sequence

<400> 98

Met Gly Ala Arg Ala Ser Ile Leu Arg Gly Gly Lys Leu Asp Thr Trp
1 5 10 15

Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys His Tyr Met Leu Lys
20 25 30

His Leu Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Leu Asn Pro
35 40 45

Gly Leu Leu Glu Thr Ser Glu Gly Cys Lys Gln Ile Met Lys Gln Leu
50 55 60

Gln Pro Ala Leu Gln Thr Gly Thr Glu Glu Leu Arg Ser Leu Tyr Asn
65 70 75 80

Thr Val Ala Thr Leu Tyr Cys Val His Glu Lys Ile Glu Val Arg Asp
85 90 95

Thr Lys Glu Ala Leu Asp Lys Ile Glu Glu Glu Gln Asn Lys Ser Gln
100 105 110

Gln Lys Thr Gln Gln Ala Glu Ala Ala Ala Asp Gly Lys Val Ser Gln
115 120 125

Asn Tyr Pro Ile Val Gln Asn Leu Gln Gly Gln Met Val His Gln Ala
130 135 140

Ile Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Ile Glu Glu Lys
145 150 155 160

Ala Phe Ser Pro Glu Val Ile Pro Met Phe Thr Ala Leu Ser Glu Gly
165 170 175

Ala Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly Gly His
180 185 190

Gln Ala Ala Met Gln Met Leu Lys Asp Thr Ile Asn Glu Glu Ala Ala
195 200 205

Glu Trp Asp Arg Leu His Pro Val His Ala Gly Pro Val Ala Pro Gly
210 215 220

Gln Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr Ser Thr
225 230 235 240

Leu Gln Glu Gln Ile Ala Trp Met Thr Ser Asn Pro Pro Val Pro Val
245 250 255

Gly Asp Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys Ile Val
260 265 270

Arg Met Tyr Ser Pro Val Ser Ile Leu Asp Ile Lys Gln Gly Pro Lys
275 280 285

Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Phe Lys Thr Leu Arg Ala
290 295 300

Glu Gln Ala Thr Gln Asp Val Lys Asn Trp Met Thr Asp Thr Leu Leu
305 310 315 320

Val Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Arg Ala Leu Gly
325 330 335

Pro Gly Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly Val Gly
340 345 350

Gly Pro Gly His Lys Ala Arg Val Leu Ala Glu Ala Met Ser Gln Ala
        355   360   365

Asn Asn Thr Asn Ile Met Met Gln Arg Ser Asn Phe Lys Gly Pro Lys
    370   375   380

Arg Ile Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His Ile Ala Arg
385   390   395   400

Asn Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys Gly Lys Glu
                405   410   415

Gly His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn Phe Leu Gly
            420   425   430

Lys Ile Trp Pro Ser His Lys Gly Arg Pro Gly Asn Phe Leu Gln Ser
        435   440   445

Arg Pro Glu Pro Thr Ala Pro Pro Ala Glu Ser Phe Arg Phe Glu Glu
    450   455   460

Thr Thr Pro Ala Pro Lys Gln Glu Pro Lys Asp Arg Glu Pro Leu Thr
465   470   475   480

Ser Leu Lys Ser Leu Phe Gly Ser Asp Pro Leu Ser Gln
                485   490

<210> 99

<211> 849

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstruction of clade C gp160 prote
       in sequence

<400> 99

Met Arg Val Met Gly Ile Gln Arg Asn Cys Gln Gln Trp Trp Ile Trp
1   5   10   15

Gly Ile Leu Gly Phe Trp Met Leu Met Ile Cys Ser Val Val Gly Asn
            20   25   30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Lys
        35   40   45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu Arg Glu Val

50 55 60

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65 70 75 80

Gln Glu Met Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
85 90 95

Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
100 105 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
115 120 125

Asn Cys Thr Asn Val Asn Asn Thr Asn Asn Thr Asn Ser Thr Met Asn
130 135 140

Gly Glu Met Lys Asn Cys Ser Phe Asn Ile Thr Thr Glu Ile Arg Asp
145 150 155 160

Lys Lys Lys Lys Glu Tyr Ala Leu Phe Tyr Arg Leu Asp Ile Val Pro
165 170 175

Leu Asn Glu Asn Asn Asn Asn Thr Ser Glu Tyr Arg Leu Ile Asn Cys
180 185 190

Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro
195 200 205

Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys
210 215 220

Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Lys Asn Val Ser Thr
225 230 235 240

Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu
245 250 255

Leu Asn Gly Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn
260 265 270

Leu Thr Asn Asn Ala Lys Thr Ile Ile Val Gln Leu Asn Glu Ser Val
275 280 285

Glu Ile Val Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Met Arg
290 295 300

Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp
305 310 315 320

Ile Arg Gln Ala His Cys Asn Ile Ser Gly Arg Glu Trp Asn Asn Thr

325 330 335

Leu Gln Gln Val Ala Glu Lys Leu Arg Lys His Phe Pro Asn Lys Thr
340 345 350

Ile Lys Phe Ala Pro Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His
355 360 365

Ser Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Lys Leu
370 375 380

Phe Asn Ser Thr Tyr Asn Ser Thr Asn Ser Thr Asn Ser Thr Ile Thr
385 390 395 400

Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Gly Val Gly
405 410 415

Gln Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Lys Ser
420 425 430

Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Lys Asn Glu Thr
435 440 445

Asn Glu Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn
450 455 460

Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Ile Lys Pro Leu
465 470 475 480

Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Glu Arg Glu Lys
485 490 495

Arg Ala Val Gly Leu Gly Ala Val Phe Leu Gly Phe Leu Gly Ala Ala
500 505 510

Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln Ala Arg
515 520 525

Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu Arg Ala
530 535 540

Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys
545 550 555 560

Gln Leu Gln Ala Arg Val Leu Ala Met Glu Arg Tyr Leu Lys Asp Gln
565 570 575

Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr
580 585 590

Ala Val Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser Gln Asp Asp Ile

595 600 605

Trp Asp Asn Met Thr Trp Met Glu Trp Asp Arg Glu Ile Asn Asn Tyr
610 615 620

Thr Asp Thr Ile Tyr Arg Leu Leu Glu Glu Ser Gln Asn Gln Gln Glu
625 630 635 640

Lys Asn Glu Gln Asp Leu Leu Ala Leu Asp Ser Trp Glu Asn Leu Trp
645 650 655

Asn Trp Phe Asp Ile Ser Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile
660 665 670

Met Ile Val Gly Gly Leu Ile Gly Leu Arg Ile Ile Phe Ala Val Leu
675 680 685

Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln
690 695 700

Thr Leu Thr Pro Asn Pro Arg Gly Pro Asp Arg Leu Glu Arg Ile Glu
705 710 715 720

Glu Glu Gly Gly Glu Gln Asp Arg Asp Arg Ser Ile Arg Leu Val Ser
725 730 735

Gly Phe Leu Ala Leu Ala Trp Asp Asp Leu Arg Ser Leu Cys Leu Phe
740 745 750

Ser Tyr His Arg Leu Arg Asp Phe Ile Leu Ile Ala Ala Arg Thr Val
755 760 765

Glu Leu Leu Gly Arg Ser Ser Leu Arg Gly Leu Gln Arg Gly Trp Glu
770 775 780

Ala Leu Lys Tyr Leu Gly Ser Leu Val Gln Tyr Trp Gly Gln Glu Leu
785 790 795 800

Lys Lys Ser Ala Ile Ser Leu Leu Asp Thr Ile Ala Ile Ala Val Ala
805 810 815

Glu Gly Thr Asp Arg Ile Ile Glu Val Val Gln Arg Ala Cys Arg Ala
820 825 830

Ile Leu Asn Ile Pro Arg Arg Ile Arg Gln Gly Phe Glu Ala Ala Leu
835 840 845

Gln

<210> 100

<211> 849

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares of center of tree reconstructions of clade C gp160
       protein sequence

<400> 100

Met Arg Val Arg Gly Ile Leu Arg Asn Cys Gln Gln Trp Trp Ile Trp
1   5   10   15

Gly Ile Leu Gly Phe Trp Met Leu Met Ile Cys Asn Val Val Gly Asn
            20   25   30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Lys
        35   40   45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu Lys Glu Val
    50   55   60

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65   70   75   80

Gln Glu Met Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85   90   95

Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
            100   105   110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115   120   125

Asn Cys Ser Asn Val Asn Ala Thr Asn Thr Thr Asn Asn Thr Met Lys
    130   135   140

Gly Glu Ile Lys Asn Cys Ser Phe Asn Ala Thr Thr Glu Ile Arg Asp
145   150   155   160

Lys Lys Gln Lys Val Tyr Ala Leu Phe Tyr Arg Leu Asp Ile Val Pro
                165   170   175

Leu Asn Glu Asn Asn Ser Asn Ser Ser Glu Tyr Arg Leu Ile Asn Cys
            180   185   190

Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro
        195   200   205

Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys
    210   215   220

Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr
225   230   235   240

Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu
                245   250   255

Leu Asn Gly Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn
            260   265   270

Leu Thr Asn Asn Val Lys Thr Ile Ile Val His Leu Asn Glu Ser Val
        275   280   285

Glu Ile Val Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg
    290   295   300

Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp
305   310   315   320

Ile Arg Gln Ala His Cys Asn Ile Ser Glu Glu Glu Trp Asn Lys Thr
                325   330   335

Leu Gln Arg Val Gly Lys Lys Leu Glu Glu His Phe Pro Asn Lys Thr
            340   345   350

Ile Lys Phe Glu Pro Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His
        355   360   365

Ser Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Lys Leu
    370   375   380

Phe Asn Ser Thr Tyr Asn Gly Thr Asn Ser Thr Asn Thr Thr Ile Thr
385   390   395   400

Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly
                405   410   415

Arg Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Lys Ser
            420   425   430

Asn Ile Thr Gly Leu Leu Leu Val Arg Asp Gly Gly Lys Asn Asn Thr
        435   440   445

Asn Asn Thr Glu Ile Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn
    450   455   460

Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Ile Lys Pro Leu
465   470   475   480

Gly Ile Ala Pro Thr Lys Ala Lys Arg Arg Val Val Glu Arg Glu Lys
485 490 495

Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu Gly Ala Ala
500 505 510

Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln Ala Arg
515 520 525

Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu Arg Ala
530 535 540

Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys
545 550 555 560

Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Lys Asp Gln
565 570 575

Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr
580 585 590

Ala Val Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser Gln Glu Asp Ile
595 600 605

Trp Asp Asn Met Thr Trp Met Gln Trp Asp Arg Glu Ile Ser Asn Tyr
610 615 620

Thr Asp Thr Ile Tyr Arg Leu Leu Glu Asp Ser Gln Asn Gln Gln Glu
625 630 635 640

Gln Asn Glu Lys Asp Leu Leu Ala Leu Asp Ser Trp Lys Asn Leu Trp
645 650 655

Asn Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile
660 665 670

Met Ile Val Gly Gly Leu Ile Gly Leu Arg Ile Ile Phe Ala Val Leu
675 680 685

Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln
690 695 700

Thr Leu Thr Pro Asn Pro Arg Gly Pro Asp Arg Leu Gly Arg Ile Glu
705 710 715 720

Glu Glu Gly Gly Glu Gln Asp Arg Asp Arg Ser Ile Arg Leu Val Ser
725 730 735

Gly Phe Leu Ala Leu Ala Trp Asp Asp Leu Arg Ser Leu Cys Leu Phe
740 745 750

Ser Tyr His Arg Leu Arg Asp Phe Ile Leu Val Ala Ala Arg Ala Val
755 760 765

Glu Leu Leu Gly Arg Ser Ser Leu Arg Gly Leu Gln Arg Gly Trp Glu
770 775 780

Ala Leu Lys Tyr Leu Gly Ser Leu Val Gln Tyr Trp Gly Leu Glu Leu
785 790 795 800

Lys Lys Ser Ala Ile Ser Leu Leu Asp Thr Ile Ala Ile Ala Val Ala
805 810 815

Glu Gly Thr Asp Arg Ile Ile Glu Leu Ile Gln Arg Ile Cys Arg Ala
820 825 830

Ile Arg Asn Ile Pro Arg Arg Ile Arg Gln Gly Phe Glu Ala Ala Leu
835 840 845

Gln

<210> 101

<211> 849

<212> PRT

<213> Artificial sequence

<220>

<223> Minimum of means of center of tree reconstructions for clade C gp 160 protein sequence

<400> 101

Met Arg Val Arg Gly Ile Leu Arg Asn Cys Gln Gln Trp Trp Ile Trp
1 5 10 15

Gly Ile Leu Gly Phe Trp Met Leu Met Ile Cys Asn Val Val Gly Asn
20 25 30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Lys
35 40 45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Glu Lys Glu Val
50 55 60

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65 70 75 80

Gln Glu Met Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
85 90 95

Asn Asp Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
100 105 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
115 120 125

Asn Cys Ser Asn Val Asn Thr Thr Asn Thr Thr Asn Asn Thr Met Lys
130 135 140

Gly Glu Ile Lys Asn Cys Ser Phe Asn Val Thr Thr Glu Leu Arg Asp
145 150 155 160

Lys Lys Lys Lys Glu Tyr Ala Leu Phe Tyr Arg Leu Asp Ile Val Pro
165 170 175

Leu Asn Glu Asn Asn Asn Asn Ser Ser Glu Tyr Arg Leu Ile Asn Cys
180 185 190

Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Asp Pro
195 200 205

Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Tyr Ala Ile Leu Lys Cys
210 215 220

Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Asn Asn Val Ser Thr
225 230 235 240

Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu
245 250 255

Leu Asn Gly Ser Leu Ala Glu Glu Glu Ile Ile Ile Arg Ser Glu Asn
260 265 270

Leu Thr Asn Asn Ala Lys Thr Ile Ile Val His Leu Asn Glu Ser Val
275 280 285

Glu Ile Val Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg
290 295 300

Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Gly Asp Ile Ile Gly Asp
305 310 315 320

Ile Arg Gln Ala His Cys Asn Ile Ser Glu Glu Glu Trp Asn Lys Thr
325 330 335

Leu Gln Arg Val Gly Lys Lys Leu Glu Glu His Phe Pro Asn Lys Thr
340 345 350

Ile Lys Phe Glu Pro Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His
355 360 365

Ser Phe Asn Cys Arg Gly Glu Phe Phe Tyr Cys Asn Thr Ser Lys Leu
    370   375   380

Phe Asn Ser Thr Tyr Asn Gly Thr Asn Ser Thr Asn Ser Thr Ile Thr
385   390   395   400

Leu Gln Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly
                405   410   415

Arg Ala Met Tyr Ala Pro Pro Ile Ala Gly Asn Ile Thr Cys Lys Ser
            420   425   430

Asn Ile Thr Gly Leu Leu Leu Val Arg Asp Gly Gly Lys Asn Asp Thr
        435   440   445

Asn Asp Thr Glu Ile Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn
    450   455   460

Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Glu Ile Lys Pro Leu
465   470   475   480

Gly Ile Ala Pro Thr Lys Ala Lys Arg Arg Val Val Glu Arg Glu Lys
                485   490   495

Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu Gly Ala Ala
            500   505   510

Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln Ala Arg
        515   520   525

Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn Leu Leu Arg Ala
    530   535   540

Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys
545   550   555   560

Gln Leu Gln Thr Arg Val Leu Ala Ile Glu Arg Tyr Leu Lys Asp Gln
                565   570   575

Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr
            580   585   590

Ala Val Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser Gln Glu Asp Ile
        595   600   605

Trp Asp Asn Met Thr Trp Met Gln Trp Asp Arg Glu Ile Ser Asn Tyr
    610   615   620

Thr Asp Thr Ile Tyr Arg Leu Leu Glu Asp Ser Gln Asn Gln Gln Glu
625   630   635   640

Gln Asn Glu Lys Asp Leu Leu Ala Leu Asp Ser Trp Lys Asn Leu Trp
645 650 655

Asn Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile
660 665 670

Met Ile Val Gly Gly Leu Ile Gly Leu Arg Ile Ile Phe Ala Val Leu
675 680 685

Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln
690 695 700

Thr Leu Thr Pro Asn Pro Arg Gly Pro Asp Arg Leu Gly Arg Ile Glu
705 710 715 720

Glu Glu Gly Gly Glu Gln Asp Arg Asp Arg Ser Ile Arg Leu Val Ser
725 730 735

Gly Phe Leu Ala Leu Ala Trp Asp Asp Leu Arg Ser Leu Cys Leu Phe
740 745 750

Ser Tyr His Arg Leu Arg Asp Phe Ile Leu Val Ala Ala Arg Ala Val
755 760 765

Glu Leu Leu Gly Arg Ser Ser Leu Arg Gly Leu Gln Arg Gly Trp Glu
770 775 780

Ala Leu Lys Tyr Leu Gly Ser Leu Val Gln Tyr Trp Gly Leu Glu Leu
785 790 795 800

Lys Lys Ser Ala Ile Ser Leu Leu Asp Thr Ile Ala Ile Ala Val Ala
805 810 815

Glu Gly Thr Asp Arg Ile Ile Glu Leu Ile Gln Arg Ile Cys Arg Ala
820 825 830

Ile Arg Asn Ile Pro Arg Arg Ile Arg Gln Gly Phe Glu Ala Ala Leu
835 840 845

Gln

<210> 102

<211> 206

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor of reconstructions of clade C nef pro
tein sequences

<400> 102

Met Gly Gly Lys Trp Ser Lys Ser Ser Ile Val Gly Trp Pro Ala Val
1   5   10   15

Arg Glu Arg Ile Arg Arg Thr Ala Pro Ala Ala Glu Gly Val Gly Ala
            20   25   30

Ala Ser Gln Asp Leu Asp Lys His Gly Ala Leu Thr Ser Ser Asn Thr
        35   40   45

Ala Ala Thr Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Glu Glu
    50   55   60

Glu Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr
65   70   75   80

Tyr Lys Gly Ala Val Asp Leu Ser Phe Phe Leu Lys Glu Lys Gly Gly
                85   90   95

Leu Glu Gly Leu Ile Tyr Ser Lys Lys Arg Gln Glu Ile Leu Asp Leu
            100   105   110

Trp Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr
        115   120   125

Pro Gly Pro Gly Ile Arg Phe Pro Leu Thr Phe Gly Trp Cys Phe Lys
    130   135   140

Leu Val Pro Val Asp Pro Arg Glu Val Glu Glu Ala Asn Glu Gly Glu
145   150   155   160

Asn Asn Cys Leu Leu His Pro Met Ser Gln His Gly Met Glu Asp Glu
                165   170   175

Asp Arg Glu Val Leu Lys Trp Lys Phe Asp Ser His Leu Ala Arg Arg
            180   185   190

His Met Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys
        195   200   205

<210> 103

<211> 207

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares and minimum of means of center of tree reconstructi
       ons of clade C nef protein sequence

<400> 103

```
Met Gly Gly Lys Trp Ser Lys Ser Ser Ile Val Gly Trp Pro Ala Val
1               5                   10                  15

Arg Glu Arg Ile Arg Arg Thr Glu Pro Ala Ala Glu Gly Val Gly Ala
            20                  25                  30

Ala Ser Gln Asp Leu Asp Lys His Gly Ala Leu Thr Ser Ser Asn Thr
        35                  40                  45

Ala Ala Asn Asn Ala Asp Cys Ala Trp Leu Glu Ala Gln Glu Glu Glu
    50                  55                  60

Glu Glu Val Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met
65                  70                  75                  80

Thr Tyr Lys Gly Ala Phe Asp Leu Ser Phe Phe Leu Lys Glu Lys Gly
                85                  90                  95

Gly Leu Glu Gly Leu Ile Tyr Ser Lys Lys Arg Gln Glu Ile Leu Asp
            100                 105                 110

Leu Trp Val Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr
        115                 120                 125

Thr Pro Gly Pro Gly Val Arg Tyr Pro Leu Thr Phe Gly Trp Cys Phe
    130                 135                 140

Lys Leu Val Pro Val Asp Pro Arg Glu Val Glu Glu Ala Asn Glu Gly
145                 150                 155                 160

Glu Asn Asn Cys Leu Leu His Pro Met Ser Gln His Gly Met Glu Asp
                165                 170                 175

Glu Asp Arg Glu Val Leu Lys Trp Lys Phe Asp Ser His Leu Ala Arg
            180                 185                 190

Arg His Met Ala Arg Glu Leu His Pro Glu Tyr Tyr Lys Asp Cys
        195                 200                 205
```

<210> 104

<211> 999

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor for reconstruction of clade C pol pro
       tein sequences

<400> 104

Phe Phe Arg Glu Asn Leu Ala Phe Pro Gln Gly Glu Ala Arg Glu Phe
1   5   10   15

Pro Ser Glu Gln Thr Arg Ala Asn Ser Pro Thr Ser Arg Glu Leu Gln
            20   25   30

Val Arg Gly Asp Asn Pro Arg Ser Glu Ala Gly Ala Glu Arg Gln Gly
        35   40   45

Thr Leu Asn Phe Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Ser
    50   55   60

Ile Lys Val Gly Gly Gln Ile Lys Glu Ala Leu Leu Asp Thr Gly Ala
65   70   75   80

Asp Asp Thr Val Leu Glu Asp Ile Asn Leu Pro Gly Lys Trp Lys Pro
                85   90   95

Lys Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp
            100   105   110

Gln Ile Leu Ile Glu Ile Cys Gly Lys Lys Ala Ile Gly Thr Val Leu
        115   120   125

Val Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn Met Leu Thr Gln
    130   135   140

Leu Gly Cys Thr Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro
145   150   155   160

Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro
                165   170   175

Leu Thr Glu Glu Lys Ile Lys Ala Leu Thr Ala Ile Cys Glu Glu Met
            180   185   190

Glu Lys Glu Gly Lys Ile Thr Lys Ile Gly Pro Glu Asn Pro Tyr Asn
        195   200   205

Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys
    210   215   220

Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu
225   230   235   240

Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser
245 250 255

Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp
260 265 270

Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn
275 280 285

Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp
290 295 300

Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu
305 310 315 320

Pro Phe Arg Ala Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp
325 330 335

Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys
340 345 350

Ile Glu Glu Leu Arg Glu His Leu Leu Lys Trp Gly Phe Thr Thr Pro
355 360 365

Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu
370 375 380

Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Gln Leu Pro Glu Lys
385 390 395 400

Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn
405 410 415

Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys
420 425 430

Leu Leu Arg Gly Ala Lys Ala Leu Thr Asp Ile Val Pro Leu Thr Glu
435 440 445

Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro
450 455 460

Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile
465 470 475 480

Gln Lys Gln Gly His Asp Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro
485 490 495

Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Lys Met Arg Ser Ala His
500 505 510

Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Ala Met
        515   520   525

Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Arg Leu Pro Ile
    530   535   540

Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Asp Tyr Trp Gln Ala Thr
545   550   555   560

Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu
                565   570   575

Trp Tyr Gln Leu Glu Lys Glu Pro Ile Ala Gly Ala Glu Thr Phe Tyr
            580   585   590

Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr
        595   600   605

Val Thr Asp Lys Gly Arg Gln Lys Val Val Ser Leu Thr Glu Thr Thr
    610   615   620

Asn Gln Lys Thr Glu Leu Gln Ala Ile Gln Leu Ala Leu Gln Asp Ser
625   630   635   640

Gly Ser Glu Val Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile
                645   650   655

Ile Gln Ala Gln Pro Asp Lys Ser Glu Ser Glu Leu Val Asn Gln Ile
            660   665   670

Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu Ser Trp Val Pro
        675   680   685

Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser
    690   695   700

Ser Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln
705   710   715   720

Glu Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Glu
                725   730   735

Phe Asn Leu Pro Pro Ile Val Ala Lys Glu Ile Val Ala Ser Cys Asp
            740   745   750

Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser
        755   760   765

Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile
    770   775   780

Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile
785   790   795   800

Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Ile Leu Lys Leu Ala
                805   810   815

Gly Arg Trp Pro Val Lys Val Ile His Thr Asp Asn Gly Ser Asn Phe
            820   825   830

Thr Ser Ala Ala Val Lys Ala Ala Cys Trp Trp Ala Gly Ile Gln Gln
        835   840   845

Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser
    850   855   860

Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala
865   870   875   880

Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe
                885   890   895

Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Ile
            900   905   910

Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile
        915   920   925

Ile Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro
    930   935   940

Val Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val
945   950   955   960

Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala
                965   970   975

Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Ala Asp Cys Val
            980   985   990

Ala Gly Arg Gln Asp Glu Asp
        995

<210> 105

<211> 999

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade C pol protei
n sequence

<400> 105

Phe Phe Arg Glu Asn Leu Ala Phe Pro Gln Gly Glu Ala Arg Glu Phe
1   5   10   15

Pro Ser Glu Gln Thr Arg Ala Asn Ser Pro Thr Ser Arg Glu Leu Gln
            20   25   30

Val Arg Gly Asp Asn Pro Arg Ser Glu Ala Gly Ala Glu Arg Gln Gly
        35   40   45

Thr Leu Asn Phe Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Ser
    50   55   60

Ile Lys Val Gly Gly Gln Ile Lys Glu Ala Leu Leu Asp Thr Gly Ala
65   70   75   80

Asp Asp Thr Val Leu Glu Glu Ile Asn Leu Pro Gly Lys Trp Lys Pro
                85   90   95

Lys Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp
            100   105   110

Gln Ile Leu Ile Glu Ile Cys Gly Lys Lys Ala Ile Gly Thr Val Leu
        115   120   125

Val Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn Met Leu Thr Gln
    130   135   140

Leu Gly Cys Thr Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro
145   150   155   160

Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro
                165   170   175

Leu Thr Glu Glu Lys Ile Lys Ala Leu Thr Ala Ile Cys Glu Glu Met
            180   185   190

Glu Lys Glu Gly Lys Ile Thr Lys Ile Gly Pro Glu Asn Pro Tyr Asn
        195   200   205

Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys
    210   215   220

Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu
225   230   235   240

Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser
                245   250   255

Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp
            260   265   270

Glu Gly Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn
        275   280   285

Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp
    290   295   300

Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu
305   310   315   320

Pro Phe Arg Ala Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp
                325   330   335

Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys
            340   345   350

Ile Glu Glu Leu Arg Glu His Leu Leu Lys Trp Gly Phe Thr Thr Pro
        355   360   365

Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu
    370   375   380

Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Gln Leu Pro Glu Lys
385   390   395   400

Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn
                405   410   415

Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys
            420   425   430

Leu Leu Arg Gly Ala Lys Ala Leu Thr Asp Ile Val Pro Leu Thr Glu
        435   440   445

Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro
    450   455   460

Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile
465   470   475   480

Gln Lys Gln Gly His Asp Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro
                485   490   495

Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Lys Met Arg Thr Ala His
            500   505   510

Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Ala Met
        515   520   525

Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Arg Leu Pro Ile
    530   535   540

Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Asp Tyr Trp Gln Ala Thr
545   550   555   560

Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu
                565   570   575

Trp Tyr Gln Leu Glu Lys Glu Pro Ile Ala Gly Ala Glu Thr Phe Tyr
            580   585   590

Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Ile Gly Lys Ala Gly Tyr
        595   600   605

Val Thr Asp Arg Gly Arg Gln Lys Ile Val Ser Leu Thr Glu Thr Thr
    610   615   620

Asn Gln Lys Thr Glu Leu Gln Ala Ile Gln Leu Ala Leu Gln Asp Ser
625   630   635   640

Gly Ser Glu Val Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile
                645   650   655

Ile Gln Ala Gln Pro Asp Lys Ser Glu Ser Glu Leu Val Asn Gln Ile
            660   665   670

Ile Glu Gln Leu Ile Lys Lys Glu Arg Val Tyr Leu Ser Trp Val Pro
        675   680   685

Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser
    690   695   700

Ser Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln
705   710   715   720

Glu Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Glu
                725   730   735

Phe Asn Leu Pro Pro Ile Val Ala Lys Glu Ile Val Ala Ser Cys Asp
            740   745   750

Lys Cys Gln Leu Lys Gly Glu Ala Ile His Gly Gln Val Asp Cys Ser
        755   760   765

Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys Ile Ile
    770   775   780

Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile
785   790   795   800

Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Ile Leu Lys Leu Ala
805 810 815

Gly Arg Trp Pro Val Lys Val Ile His Thr Asp Asn Gly Ser Asn Phe
820 825 830

Thr Ser Ala Ala Val Lys Ala Ala Cys Trp Trp Ala Gly Ile Gln Gln
835 840 845

Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser
850 855 860

Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala
865 870 875 880

Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe
885 890 895

Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Ile
900 905 910

Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile
915 920 925

Ile Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro
930 935 940

Ile Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val
945 950 955 960

Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala
965 970 975

Lys Ile Ile Lys Asp Tyr Gly Lys Gln Met Ala Gly Ala Asp Cys Val
980 985 990

Ala Gly Arg Gln Asp Glu Asp
995

<210> 106

<211> 999

<212> PRT

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade C pol pro
tein sequence

<400> 106

Phe Phe Arg Glu Asn Leu Ala Phe Pro Gln Gly Glu Ala Arg Glu Phe
1   5   10   15

Pro Ser Glu Gln Thr Arg Ala Asn Ser Pro Thr Ser Arg Glu Leu Gln
            20   25   30

Val Arg Gly Asp Asn Pro Leu Ser Glu Ala Gly Ala Glu Arg Gln Gly
        35   40   45

Thr Leu Asn Phe Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr
    50   55   60

Ile Lys Val Gly Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala
65   70   75   80

Asp Asp Thr Val Leu Glu Glu Ile Asn Leu Pro Gly Lys Trp Lys Pro
                85   90   95

Lys Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp
            100   105   110

Gln Ile Leu Ile Glu Ile Cys Gly Lys Lys Ala Ile Gly Thr Val Leu
        115   120   125

Val Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn Met Leu Thr Gln
    130   135   140

Leu Gly Cys Thr Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro
145   150   155   160

Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro
                165   170   175

Leu Thr Glu Glu Lys Ile Lys Ala Leu Thr Ala Ile Cys Glu Glu Met
            180   185   190

Glu Lys Glu Gly Lys Ile Thr Lys Ile Gly Pro Glu Asn Pro Tyr Asn
        195   200   205

Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys
    210   215   220

Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu
225   230   235   240

Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser
                245   250   255

Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp
            260   265   270

Glu Gly Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn
        275   280   285

Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp
    290   295   300

Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu
305   310   315   320

Pro Phe Arg Ala Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp
                325   330   335

Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys
            340   345   350

Ile Glu Glu Leu Arg Glu His Leu Leu Lys Trp Gly Phe Thr Thr Pro
        355   360   365

Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu
    370   375   380

Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Gln Leu Pro Glu Lys
385   390   395   400

Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn
                405   410   415

Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys
            420   425   430

Leu Leu Arg Gly Ala Lys Ala Leu Thr Asp Ile Val Pro Leu Thr Glu
        435   440   445

Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro
    450   455   460

Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile
465   470   475   480

Gln Lys Gln Gly His Asp Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro
                485   490   495

Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Lys Met Arg Thr Ala His
            500   505   510

Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Ala Met
        515   520   525

Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Arg Leu Pro Ile
    530   535   540

Gln Lys Glu Thr Trp Glu Ala Trp Trp Thr Asp Tyr Trp Gln Ala Thr
545 550 555 560

Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu
565 570 575

Trp Tyr Gln Leu Glu Lys Glu Pro Ile Ala Gly Ala Glu Thr Phe Tyr
580 585 590

Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Ile Gly Lys Ala Gly Tyr
595 600 605

Val Thr Asp Arg Gly Arg Gln Lys Ile Val Ser Leu Thr Glu Thr Thr
610 615 620

Asn Gln Lys Thr Glu Leu Gln Ala Ile Gln Leu Ala Leu Gln Asp Ser
625 630 635 640

Gly Ser Glu Val Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile
645 650 655

Ile Gln Ala Gln Pro Asp Lys Ser Glu Ser Glu Leu Val Asn Gln Ile
660 665 670

Ile Glu Gln Leu Ile Lys Lys Glu Arg Val Tyr Leu Ser Trp Val Pro
675 680 685

Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser
690 695 700

Ser Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln
705 710 715 720

Glu Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Glu
725 730 735

Phe Asn Leu Pro Pro Ile Val Ala Lys Glu Ile Val Ala Ser Cys Asp
740 745 750

Lys Cys Gln Leu Lys Gly Glu Ala Ile His Gly Gln Val Asp Cys Ser
755 760 765

Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys Ile Ile
770 775 780

Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile
785 790 795 800

Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Ile Leu Lys Leu Ala
805 810 815

Gly Arg Trp Pro Val Lys Val Ile His Thr Asp Asn Gly Ser Asn Phe
            820   825   830

Thr Ser Ala Ala Val Lys Ala Ala Cys Trp Trp Ala Gly Ile Gln Gln
        835   840   845

Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser
    850   855   860

Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala
865   870   875   880

Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe
                885   890   895

Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Ile
            900   905   910

Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile
        915   920   925

Ile Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro
    930   935   940

Ile Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val
945   950   955   960

Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala
                965   970   975

Lys Ile Ile Lys Asp Tyr Gly Lys Gln Met Ala Gly Ala Asp Cys Val
            980   985   990

Ala Gly Arg Gln Asp Glu Asp
        995

<210> 107

<211> 107

<212> PRT

<213> Artificial sequence

<220>

<223> Most common recent ancestor reconstructions of clade C rev protei
       n sequence

<400> 107

Met Ala Gly Arg Ser Gly Asp Ser Asp Glu Ala Leu Leu Gln Ala Val

```
1               5                   10                  15

Arg Ile Ile Lys Ile Leu Tyr Gln Ser Asn Pro Tyr Pro Lys Pro Glu
            20                  25                  30

Gly Thr Arg Gln Ala Arg Arg Asn Arg Arg Arg Arg Trp Arg Ala Arg
        35                  40                  45

Gln Arg Gln Ile His Ser Ile Ser Glu Arg Ile Leu Ser Thr Cys Leu
    50                  55                  60

Gly Arg Pro Ala Glu Pro Val Pro Leu Gln Leu Pro Pro Leu Glu Arg
65                  70                  75                  80

Leu His Leu Asp Cys Ser Glu Asp Cys Gly Thr Ser Gly Thr Gln Gln
                85                  90                  95

Ser Gln Gly Thr Thr Glu Gly Val Gly Ser Pro
            100                 105
```

<210> 108

<211> 107

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstructions of clade C rev prote
in sequence

<400> 108

```
Met Ala Gly Arg Ser Gly Asp Ser Asp Glu Ala Leu Leu Gln Ala Val
1               5                   10                  15

Arg Ile Ile Lys Ile Leu Tyr Gln Ser Asn Pro Tyr Pro Lys Pro Glu
            20                  25                  30

Gly Thr Arg Gln Ala Arg Lys Asn Arg Arg Arg Arg Trp Arg Ala Arg
        35                  40                  45

Gln Arg Gln Ile His Ser Ile Ser Glu Arg Ile Leu Ser Thr Cys Leu
    50                  55                  60

Gly Arg Pro Ala Glu Pro Val Pro Leu Gln Leu Pro Pro Ile Glu Arg
65                  70                  75                  80

Leu His Ile Gly Asp Ser Glu Ser Ser Gly Thr Ser Gly Thr Gln Gln
                85                  90                  95
```

Ser Gln Gly Thr Thr Glu Gly Val Gly Ser Pro
100 105

<210> 109

<211> 107

<212> PRT

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstructions of clade C rev pr
otein sequence

<400> 109

Met Ala Gly Arg Ser Gly Asp Ser Asp Glu Ala Leu Leu Gln Ala Val
1 5 10 15

Arg Ile Ile Lys Ile Leu Tyr Gln Ser Asn Pro Tyr Pro Lys Pro Glu
20 25 30

Gly Thr Arg Gln Ala Arg Lys Asn Arg Arg Arg Arg Trp Arg Ala Arg
35 40 45

Gln Arg Gln Ile His Ser Ile Ser Glu Arg Ile Leu Ser Thr Cys Leu
50 55 60

Gly Arg Pro Ala Glu Pro Val Pro Phe Gln Leu Pro Pro Ile Glu Arg
65 70 75 80

Leu His Ile Gly Asp Ser Glu Ser Ser Gly Thr Ser Gly Thr Gln Gln
85 90 95

Ser Gln Gly Thr Thr Glu Gly Val Gly Ser Pro
100 105

<210> 110

<211> 101

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstructions of clade C tat protei
n sequence

<400> 110

Met Glu Pro Val Asp Pro Asn Leu Glu Pro Trp Asn His Pro Gly Ser
1 5 10 15

Gln Pro Lys Thr Ala Cys Asn Lys Cys Tyr Cys Lys Lys Cys Ser Tyr
20 25 30

His Cys Leu Val Cys Phe Leu Thr Lys Gly Leu Gly Ile Ser Tyr Gly
35 40 45

Arg Lys Lys Arg Arg Gln Arg Arg Arg Ala Pro Pro Ser Ser Glu Asp
50 55 60

His Gln Asn Pro Ile Ser Lys Gln Pro Leu Ser Gln Thr Arg Gly Asp
65 70 75 80

Pro Thr Gly Ser Glu Glu Ser Lys Lys Lys Val Glu Ser Lys Thr Glu
85 90 95

Thr Asp Pro Cys Asp
100

<210> 111

<211> 101

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares and minimum of means center of tree reconstruction of clade C tat protein sequence

<400> 111

Met Glu Pro Val Asp Pro Asn Leu Glu Pro Trp Asn His Pro Gly Ser
1 5 10 15

Gln Pro Lys Thr Pro Cys Asn Lys Cys Tyr Cys Lys His Cys Ser Tyr
20 25 30

His Cys Leu Val Cys Phe Gln Thr Lys Gly Leu Gly Ile Ser Tyr Gly
35 40 45

Arg Lys Lys Arg Arg Gln Arg Arg Ser Ala Pro Pro Ser Ser Glu Asp
50 55 60

His Gln Asn Pro Ile Ser Lys Gln Pro Leu Pro Gln Thr Arg Gly Asp
65 70 75 80

Pro Thr Gly Ser Glu Glu Ser Lys Lys Lys Val Glu Ser Lys Thr Glu
85 90 95

Thr Asp Pro Phe Asp

100

<210> 112

<211> 192

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstructions of clade C vif protei
n sequence

<400> 112

Met Glu Asn Arg Trp Gln Val Leu Ile Val Trp Gln Val Asp Arg Met
1   5   10   15

Lys Ile Arg Thr Trp Asn Ser Leu Val Lys His His Met Tyr Val Ser
            20   25   30

Arg Arg Ala Lys Gly Trp Phe Tyr Arg His His Tyr Glu Ser Arg His
        35   40   45

Pro Lys Ile Ser Ser Glu Val His Ile Pro Leu Gly Asp Ala Arg Leu
    50   55   60

Val Ile Lys Thr Tyr Trp Gly Leu His Thr Gly Glu Arg Asp Trp His
65   70   75   80

Leu Gly His Gly Val Ser Ile Glu Trp Arg Leu Arg Arg Tyr Ser Thr
                85   90   95

Gln Val Asp Pro Gly Leu Ala Asp Gln Leu Ile His Met His Tyr Phe
            100   105   110

Asp Cys Phe Ala Asp Ser Ala Ile Arg Lys Ala Ile Leu Gly His Ile
        115   120   125

Val Ser Pro Arg Cys Asp Tyr Gln Ala Gly His Asn Lys Val Gly Ser
    130   135   140

Leu Gln Tyr Leu Ala Leu Thr Ala Leu Ile Lys Pro Lys Lys Ile Lys
145   150   155   160

Pro Pro Leu Pro Ser Val Lys Lys Leu Val Glu Asp Arg Trp Asn Lys
                165   170   175

Pro Gln Lys Thr Arg Gly His Arg Gly Ser His Thr Met Asn Gly His
            180   185   190

<210> 113

<211> 192

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade C vif protei
n sequence

<400> 113

Met Glu Asn Arg Trp Gln Val Leu Ile Val Trp Gln Val Asp Arg Met
1   5   10   15

Lys Ile Arg Thr Trp Asn Ser Leu Val Lys His His Met Tyr Val Ser
            20   25   30

Arg Arg Ala Asn Gly Trp Phe Tyr Arg His His Tyr Glu Ser Arg His
        35   40   45

Pro Lys Val Ser Ser Glu Val His Ile Pro Leu Gly Asp Ala Arg Leu
    50   55   60

Val Ile Lys Thr Tyr Trp Gly Leu Gln Thr Gly Glu Arg Asp Trp His
65   70   75   80

Leu Gly His Gly Val Ser Ile Glu Trp Arg Leu Arg Arg Tyr Ser Thr
                85   90   95

Gln Val Asp Pro Gly Leu Ala Asp Gln Leu Ile His Met His Tyr Phe
            100   105   110

Asp Cys Phe Ala Asp Ser Ala Ile Arg Lys Ala Ile Leu Gly His Ile
        115   120   125

Val Ile Pro Arg Cys Asp Tyr Gln Ala Gly His Asn Lys Val Gly Ser
    130   135   140

Leu Gln Tyr Leu Ala Leu Thr Ala Leu Ile Lys Pro Lys Lys Ile Lys
145   150   155   160

Pro Pro Leu Pro Ser Val Arg Lys Leu Val Glu Asp Arg Trp Asn Lys
                165   170   175

Pro Gln Lys Thr Arg Gly Arg Arg Gly Asn His Thr Met Asn Gly His
            180   185   190

<210> 114

<211> 192

<212> PRT

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstructions of clade C vif pr
otein sequence

<400> 114

Met Glu Asn Arg Trp Gln Val Leu Ile Val Trp Gln Val Asp Arg Met
1 5 10 15

Lys Ile Arg Thr Trp Asn Ser Leu Val Lys His His Met Tyr Val Ser
20 25 30

Arg Arg Ala Asn Gly Trp Phe Tyr Arg His His Tyr Glu Ser Arg His
35 40 45

Pro Lys Val Ser Ser Glu Val His Ile Pro Leu Gly Asp Ala Arg Leu
50 55 60

Val Ile Lys Thr Tyr Trp Gly Leu His Thr Gly Glu Arg Asp Trp His
65 70 75 80

Leu Gly His Gly Val Ser Ile Glu Trp Arg Leu Arg Arg Tyr Ser Thr
85 90 95

Gln Val Asp Pro Gly Leu Ala Asp Gln Leu Ile His Met His Tyr Phe
100 105 110

Asp Cys Phe Ala Asp Ser Ala Ile Arg Lys Ala Ile Leu Gly His Ile
115 120 125

Val Ile Pro Arg Cys Asp Tyr Gln Ala Gly His Asn Lys Val Gly Ser
130 135 140

Leu Gln Tyr Leu Ala Leu Thr Ala Leu Ile Lys Pro Lys Lys Ile Lys
145 150 155 160

Pro Pro Leu Pro Ser Val Arg Lys Leu Val Glu Asp Arg Trp Asn Lys
165 170 175

Pro Gln Lys Thr Arg Gly Arg Arg Gly Asn His Thr Met Asn Gly His
180 185 190

<210> 115

<211> 96

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstructions for clade C vpr prote
in sequence

<400> 115

Met Glu Gln Ala Pro Glu Asp Gln Gly Pro Gln Arg Glu Pro Tyr Asn
1   5   10   15

Glu Trp Thr Leu Glu Leu Leu Glu Glu Leu Lys Gln Glu Ala Val Arg
            20   25   30

His Phe Pro Arg Pro Trp Leu His Ser Leu Gly Gln His Ile Tyr Glu
        35   40   45

Thr Tyr Gly Asp Thr Trp Ala Gly Val Glu Ala Ile Ile Arg Ile Leu
    50   55   60

Gln Gln Leu Leu Phe Ile His Phe Arg Ile Gly Cys Gln His Ser Arg
65   70   75   80

Ile Gly Ile Ile Arg Gln Arg Arg Ala Arg Asn Gly Ala Ser Arg Ser
                85   90   95

<210> 116

<211> 96

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares center of tree reconstruction of clade C vpr protei
n sequence

<400> 116

Met Glu Gln Ala Pro Glu Asp Gln Gly Pro Gln Arg Glu Pro Tyr Asn
1   5   10   15

Glu Trp Thr Leu Glu Ile Leu Glu Glu Leu Lys Gln Glu Ala Val Arg
            20   25   30

His Phe Pro Arg Pro Trp Leu His Ser Leu Gly Gln Tyr Ile Tyr Glu
        35   40   45

Thr Tyr Gly Asp Thr Trp Thr Gly Val Glu Ala Leu Ile Arg Ile Leu
    50   55   60

Gln Gln Leu Leu Phe Ile His Phe Arg Ile Gly Cys Gln His Ser Arg
65 70 75 80

Ile Gly Ile Leu Arg Gln Arg Arg Ala Arg Asn Gly Ala Ser Arg Ser
85 90 95

<210> 117

<211> 96

<212> PRT

<213> Artificial sequence

<220>

<223> Minimum of means center of tree reconstruction of clade C vpr pro
tein sequence

<400> 117

Met Glu Gln Ala Pro Glu Asp Gln Gly Pro Gln Arg Glu Pro Tyr Asn
1 5 10 15

Glu Trp Thr Leu Glu Leu Leu Glu Glu Leu Lys Gln Glu Ala Val Arg
20 25 30

His Phe Pro Arg Pro Trp Leu His Ser Leu Gly Gln His Ile Tyr Glu
35 40 45

Thr Tyr Gly Asp Thr Trp Thr Gly Val Glu Ala Leu Ile Arg Ile Leu
50 55 60

Gln Gln Leu Leu Phe Ile His Phe Arg Ile Gly Cys Gln His Ser Arg
65 70 75 80

Ile Gly Ile Met Arg Gln Arg Arg Ala Arg Asn Gly Ala Ser Arg Ser
85 90 95

<210> 118

<211> 86

<212> PRT

<213> Artificial sequence

<220>

<223> Most recent common ancestor reconstructions for clade C vpu prote
in sequence

<400> 118

Met Leu Asp Leu Ile Ala Arg Val Asp Tyr Arg Leu Gly Val Gly Ala
1 5 10 15

Leu Ile Val Ala Leu Ile Ile Ala Ile Val Val Trp Thr Ile Val Tyr
20 25 30

Ile Glu Tyr Arg Lys Leu Val Arg Gln Arg Lys Ile Asp Trp Leu Ile
35 40 45

Lys Arg Ile Arg Glu Arg Ala Glu Asp Ser Gly Asn Glu Ser Asp Gly
50 55 60

Asp Thr Glu Glu Leu Ser Thr Leu Val Asp Met Gly His Leu Arg Leu
65 70 75 80

Leu Asp Val Asn Asp Leu
85

<210> 119

<211> 86

<212> PRT

<213> Artificial sequence

<220>

<223> Least squares and minimum of means center of tree reconstruction of clade C vpu protein sequence

<400> 119

Met Leu Asp Leu Leu Ala Arg Val Asp Tyr Arg Leu Gly Val Gly Ala
1 5 10 15

Leu Ile Val Ala Leu Ile Ile Ala Ile Val Val Trp Thr Ile Val Tyr
20 25 30

Ile Glu Tyr Arg Lys Leu Leu Arg Gln Arg Lys Ile Asp Trp Leu Ile
35 40 45

Lys Arg Ile Arg Glu Arg Ala Glu Asp Ser Gly Asn Glu Ser Glu Gly
50 55 60

Asp Thr Glu Glu Leu Ser Thr Met Val Asp Met Gly His Leu Arg Leu
65 70 75 80

Leu Asp Val Asn Asp Leu
85

<210> 120

<211> 376

<212> PRT

<213> Artificial sequence

<220>

<223> Deduced ancestor env protein sequence

<400> 120

Asn Lys Ser Glu Thr Asp Arg Trp Gly Leu Thr Lys Ser Ser Thr Thr
1   5   10   15

Thr Thr Thr Ala Ala Pro Thr Ser Ala Pro Val Ser Glu Lys Ile Asp
            20   25   30

Met Val Asn Glu Thr Ser Ser Cys Ile Ala Gln Asn Asn Cys Thr Gly
        35   40   45

Leu Glu Gln Glu Gln Met Ile Ser Cys Lys Phe Asn Met Thr Gly Leu
    50   55   60

Lys Arg Asp Lys Thr Lys Glu Tyr Asn Glu Thr Trp Tyr Ser Thr Asp
65   70   75   80

Leu Val Cys Glu Gln Gly Asn Ser Thr Asp Asn Glu Ser Arg Cys Tyr
                85   90   95

Met Asn His Cys Asn Thr Ser Val Ile Gln Glu Ser Cys Asp Lys His
            100   105   110

Tyr Trp Asp Thr Ile Arg Phe Arg Tyr Cys Ala Pro Pro Gly Tyr Ala
        115   120   125

Leu Leu Arg Cys Asn Asp Thr Asn Tyr Ser Gly Phe Met Pro Lys Cys
    130   135   140

Ser Lys Val Val Val Ser Ser Cys Thr Arg Met Met Glu Thr Gln Thr
145   150   155   160

Ser Thr Trp Phe Gly Phe Asn Gly Thr Arg Ala Glu Asn Arg Thr Tyr
                165   170   175

Ile Tyr Trp His Gly Arg Asp Asn Arg Thr Ile Ile Ser Leu Asn Lys
            180   185   190

Tyr Tyr Asn Leu Thr Met Lys Cys Arg Arg Pro Gly Asn Lys Thr Val
        195   200   205

Leu Pro Val Thr Ile Met Ser Gly Leu Val Phe His Ser Gln Pro Ile
    210   215   220

Asn Asp Arg Pro Lys Gln Ala Trp Cys Trp Phe Gly Gly Lys Trp Lys

225 230 235 240

Asp Ala Ile Lys Glu Val Lys Gln Thr Ile Val Lys His Pro Arg Tyr
245 250 255

Thr Gly Thr Asn Asn Thr Asp Lys Ile Asn Leu Thr Ala Pro Gly Gly
260 265 270

Gly Asp Pro Glu Val Thr Phe Met Trp Thr Asn Cys Arg Gly Glu Phe
275 280 285

Leu Tyr Cys Lys Met Asn Trp Phe Leu Asn Trp Val Glu Asp Arg Asp
290 295 300

Val Thr Thr Gln Arg Pro Lys Glu Arg His Arg Arg Asn Tyr Val Pro
305 310 315 320

Cys His Ile Arg Gln Ile Ile Asn Thr Trp His Lys Val Gly Lys Asn
325 330 335

Val Tyr Leu Pro Pro Arg Glu Gly Asp Leu Thr Cys Asn Ser Thr Val
340 345 350

Thr Ser Leu Ile Ala Asn Ile Asp Trp Thr Asp Gly Asn Gln Thr Asn
355 360 365

Ile Thr Met Ser Ala Glu Val Ala
370 375

<210> 121

<211> 883

<212> PRT

<213> Artificial sequence

<220>

<223> Deduced ancestor env protein sequence

<400> 121

Met Arg Val Lys Gly Ile Arg Lys Asn Tyr Gln His Leu Trp Arg Trp
1 5 10 15

Gly Thr Met Leu Leu Gly Met Leu Met Ile Cys Ser Ala Ala Glu Lys
20 25 30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr
35 40 45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val

50 55 60

His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65 70 75 80

Gln Glu Val Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
85 90 95

Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
100 105 110

Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
115 120 125

Asn Cys Thr Asp Asp Leu Arg Thr Asn Ala Thr Asn Thr Thr Asn Ser
130 135 140

Ser Ala Thr Thr Asn Thr Thr Ser Ser Gly Gly Gly Thr Met Glu Gly
145 150 155 160

Glu Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn Val Thr Thr Ser Ile
165 170 175

Arg Asp Lys Met Gln Lys Glu Tyr Ala Leu Phe Tyr Lys Leu Asp Val
180 185 190

Val Pro Ile Asp Asn Asp Asn Asn Asn Thr Asn Asn Asn Thr Ser Tyr
195 200 205

Arg Leu Ile Asn Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro Lys
210 215 220

Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Thr Pro Ala Gly Phe
225 230 235 240

Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn Gly Thr Gly Pro Cys
245 250 255

Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val Val
260 265 270

Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Val Val
275 280 285

Ile Arg Ser Glu Asn Phe Thr Asp Asn Ala Lys Thr Ile Ile Val Gln
290 295 300

Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr
305 310 315 320

Arg Lys Ser Ile Pro Ile Gly Pro Gly Arg Ala Leu Tyr Ala Thr Gly

325 330 335

Lys Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Leu Ser Arg Ala
340 345 350

Lys Trp Asn Asn Thr Leu Lys Gln Ile Val Thr Lys Leu Arg Glu Gln
355 360 365

Phe Gly Asn Asn Lys Thr Thr Ile Val Phe Asn Gln Ser Ser Gly Gly
370 375 380

Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu Phe Phe
385 390 395 400

Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp His Phe Asn Gly
405 410 415

Thr Trp Gly Asn Asn Asn Thr Glu Arg Ser Asn Asn Ala Ala Asp Asp
420 425 430

Asn Asp Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met
435 440 445

Trp Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro Ile Ser Gly Gln
450 455 460

Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly
465 470 475 480

Gly Asn Asn Glu Asn Thr Asn Asn Thr Asp Thr Glu Ile Phe Arg Pro
485 490 495

Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr
500 505 510

Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys
515 520 525

Arg Arg Val Val Gln Arg Glu Lys Arg Ala Val Gly Met Leu Gly Ala
530 535 540

Met Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala
545 550 555 560

Ser Met Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val
565 570 575

Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu
580 585 590

Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg Val Leu

595 600 605

Ala Val Glu Arg Tyr Leu Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly
610 615 620

Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro Trp Asn Ala Ser
625 630 635 640

Trp Ser Asn Lys Ser Leu Asp Lys Ile Trp Asn Asn Met Thr Trp Met
645 650 655

Glu Trp Glu Arg Glu Ile Asp Asn Tyr Thr Gly Leu Ile Tyr Thr Leu
660 665 670

Ile Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu
675 680 685

Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp Phe Asp Ile Thr Asn
690 695 700

Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met Ile Val Gly Gly Leu Val
705 710 715 720

Gly Leu Arg Ile Val Phe Ala Val Leu Ser Ile Val Asn Arg Val Arg
725 730 735

Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr Arg Leu Pro Ala Pro Arg
740 745 750

Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu Glu Gly Gly Glu Arg Asp
755 760 765

Arg Asp Arg Ser Gly Arg Leu Val Asn Gly Phe Leu Ala Leu Ile Trp
770 775 780

Asp Asp Leu Arg Ser Leu Cys Leu Phe Ser Tyr His Arg Leu Arg Asp
785 790 795 800

Leu Leu Leu Ile Val Ala Arg Ile Val Glu Leu Leu Gly Arg Arg Gly
805 810 815

Trp Glu Ala Leu Lys Tyr Trp Trp Asn Leu Leu Gln Tyr Trp Ser Gln
820 825 830

Glu Leu Lys Asn Ser Ala Val Ser Leu Leu Asn Ala Thr Ala Ile Ala
835 840 845

Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val Val Gln Arg Ala Cys
850 855 860

Arg Ala Ile Leu His Ile Pro Arg Arg Ile Arg Gln Gly Leu Glu Arg

865   870   875   880

Ala Leu Leu

<210> 122

<211> 25

<212> DNA

<213> Artificial sequence

<220>

<223> First round primer UP-3

<400> 122
agactgcaga tgtgaagagg tacac   25

<210> 123

<211> 22

<212> DNA

<213> Artificial sequence

<220>

<223> First round primer PEXTM6

<400> 123
ggatctggta tgctcatagc aa   22

<210> 124

<211> 25

<212> DNA

<213> Artificial sequence

<220>

<223> Second round primers PEXTM7

<400> 124
gatactgcag caacagcaac agctg   25

<210> 125

<211> 24

<212> DNA

<213> Artificial sequence

<220>

<223> Second round primer UP-5

<400> 125
gcaaagcttc tctggttggc agtg   24